A simple estimation method and finite-sample inference for a stochastic volatility model

Jean-Marie Dufour* Pascale Valéry†

22nd January 2004

ABSTRACT

The aim of the paper is to fulfill the gap for testing hypotheses on parameters of the log-normal stochastic volatility model, more precisely, to propose finite sample exact tests in the sense that the tests have correct levels in small samples. To do this, we examine method-of-moments-based tests and we provide explicit expressions for all the moments and the estimators which simplifies highly the test procedures. We then state the asymptotic distribution of the estimator as well as that of the proposed test statistics for testing the null hypothesis of no persistence in the volatility. We then compare in a study of level and power the standard asymptotic techniques to the technique of Monte Carlo tests which is valid in finite samples and allow for test statistics whose null distribution may depend on nuisance parameters. In particular maximized Monte Carlo tests (MMC) introduced by Dufour (1995) have the exact level in finite samples when the p-value function is maximized over the entire set of nuisance parameters. In contrast to MMC tests which are highly computer intensive, simplified (asymptotically justified) approximate versions of Monte Carlo tests provide a halfway solution which achieves to control the level of the tests while alleviating the use of computers.

Key words: exact tests; Monte Carlo tests; $C(\alpha)$-tests; finite sample tests; stochastic volatility; Method-of-moments .
JEL classification: C1, C13, C12, C32, C15

*CIREQ,CIRANO,Université de Montreal,email: jean.marie.dufour@umontreal.ca
†CIREQ,CIRANO,Université de Montreal,email: valeryp@magellan.umontreal.ca

Contents

List of Definitions, Propositions and Theorems

List of Tables

List of Figures

1. Introduction

Evaluating the likelihood function of ARCH models is relatively easy compared to Stochastic Volatility models (SV) for which it is impossible to get an explicit closed-form expression for the likelihood function [see Shephard (1996), Mahieu and Schotman (1998)]. This is a generic feature common to almost all nonlinear latent variable models due to the curse of the high dimensionality of the integral appearing in the likelihood function of the stochastic volatility model. This is the reason why econometricians were reluctant to use this kind of models in their applications for a long time since in this setting, maximum likelihood methods are computationnaly intensive. But recently progress has been made regarding the estimation of nonlinear latent variable models in general and stochastic volatility models in particular. There mainly exists three types of methods, namely, the quasi-exact methods, simulation-based-estimation methods and the bayesian methods. Thus, we can mention the Quasi Maximum Likelihood (QML) approach suggested by Nelson (1988) and Harvey, Ruiz and Shephard (1994), Ruiz (1994), and a Generalized Method of Moments(GMM) procedure proposed by Melino and Turnbull (1990). On the other hand, increased computer power has made simulation-based estimation methods more attractive among which we can mention the Simulated Method of Moments (SMM) proposed by Duffie and Singleton (1993), the indirect inference approach of Gouriéroux, Monfort and Renault (1993) and the moment matching methods of Gallant and Tauchen (1994). But computer intensive Markov Chain Monte Carlo methods applied to SV models by Jacquier, Polson and Rossi (1994) and Kim and Shephard (1994), Kim, Shephard and Chib (1998), Wong(2002a,2002b) and simulation-based Maximum Likelihood (SML) method proposed by Danielsson and Richard (1993), Danielsson (1994), are the most efficient methods to estimate this kind of models. In particular, Danielsson (1994), Danielsson and Richard (1993) developp an importance sampling technique to estimate the integral appearing in the likelihood function of the SV model. In a Bayesian setting, Jacquier, Polson and Rossi (1994), Kim, Shephard and Chib (1998) combine a Gibbs sampler with the Metropolis-Hastings algorithm to obtain the marginal posterior densities of the parameters of the SV model.

This paper has two contributions. The first one is to propose a simple estimation method for a log-normal stochastic volatility model with an autoregressive mean part. The second contribution which appears to be the most important one, is to provide inference techniques for this model.

Indeed, the standard form as set forth, for instance, in Harvey, Ruiz, and Shephard (1994), Jacquier, Polson, and Rossi (1994), Danielsson (1994), takes the form of an autoregression whose innovations are scaled by an unobservable volatility process, usually distributed as a lognormal autoregression but other distributions (Student, mixture of normal distributions) can be considered [see Kim, Shephard and Chib

(1998), Mahieu and Schotman (1998), Wong (2002a,2002b)]. The stochastic volatility specification we have chosen here comes from Gallant, Hsieh, Tauchen (1995), Tauchen (1996). Whereas all the authors quoted above, mainly focus on estimation procedures for the stochastic volatility model, often preoccupied by efficiency considerations [e.g. bayesian methods, Efficient Method of Moments], our paper unlike the others is mostly motivated by inference techniques applicable to the stochastic volatility model. Our concern for inference, in particular for simulation-based inference such as the technique of Monte Carlo tests introduced by Dwass (1957) for permutation tests, and later extended by Barnard (1963) and Birnbaum (1974), justifies an estimation method easy to implement. Thus, the estimation method used in this paper is mainly a method of moments in two steps which coincides with the GMM procedure in the particular case that the autoregressive mean part vanishes. Econometricians previously quoted mainly focused on efficient estimation procedures to estimate the SV model, they mostly examine specification tests such as the χ^2 tests for goodness of fit in Andersen and Sorensen (1996), Andersen, Chung and Sorensen (1999), specification tests with diagnostics in Gallant, Hsieh and Tauchen (1995), χ^2 specification tests through Indirect Inference criterion in Monfardini (1997), or likelihood ratio tests statistics for comparative fit in Kim, Shephard and Chib (1998). As a consequence, inference techniques for testing hypotheses on parameters remained underdeveloped, apart from standard t tests for individual parameters in Andersen and Sorensen (1996), in Andersen, Chung and Sorensen (1999) often performed with size distorsions.

In this setting, the aim of the paper is to fulfill the gap for testing hypotheses on parameters of the SV model, more precisely, to propose finite sample exact tests in the sense that the tests have correct levels in small samples. To do this, we examine method-of-moments-based tests. We extend the first moments of the volatility process originally stated by Jacquier, Polson and Rossi (1994) by providing general expressions for them and further provide analytic formulas for the estimators which simplifies highly the test procedures. We then state the asymptotic distribution of the estimator as well as that of the proposed test statistics for testing the null hypothesis of no persistence in the volatility. We then compare in a study of level and power the standard asymptotic techniques to the technique of Monte Carlo tests which is valid in finite samples and allow for test statistics whose null distribution may depend on nuisance parameters. In particular maximized Monte Carlo tests (MMC) introduced by Dufour (1995) have the exact level in finite samples when the p-value function is maximized over the entire set of nuisance parameters. In contrast to MMC tests which are highly computer intensive, simplified (asymptotically justified) approximate versions of Monte Carlo tests provide a halfway solution which achieves to control the level of the tests while alleviating the use of computers.

The paper is organized as follows. The second section sets the framework and the assumptions underlying the model. In the third section, we expose the estimation procedure used as well as the distributional results obtained for our estimator. Hypothesis testing is examined in the fourth section and the distribution of the test statistics is stated. The fifth section explicits the technique of Monte Carlo tests. The sixth section presents the data used in the empirical application while implementation results are discussed in the seventh section. All proofs are gathered in the appendix.

2. Framework

The basic form of the stochastic volatility model we study here for y_t comes from Gallant, Hsieh, Tauchen (1995). Let y_t denote the first difference over a short time interval, a day for instance, of the log-price of a financial asset traded on securities markets.

Assumption 2.1 *The process $\{y_t,\ t \in \mathbb{N}\}$ follows a stochastic volatility model of the type:*

$$y_t - \mu_y = \sum_{j=1}^{L_y} c_j(y_{t-j} - \mu_y) + exp(w_t/2) r_y z_t \ , \tag{2.1}$$

$$w_t - \mu_w = \sum_{j=1}^{L_w} a_j(w_{t-j} - \mu_w) + r_w v_t \ , \tag{2.2}$$

where μ_y, $\{c_j\}_{j=1}^{L_y}$, r_y, μ_w, $\{a_j\}_{j=1}^{L_w}$ and r_w are the parameters of the two equations, called the mean and volatility equations respectively. $s_t = (y_t, w_t)'$ is initialized from its stationary distribution.

The lag lengths of the autoregressive specifications used in the literature are typically short, e.g. $L_w = 1$, and $L_y = 1$, or $L_y = 0$ [see e.g. Andersen and Sorensen (1996), Andersen, Chung and Sorensen (1999) Gallant, Hsieh, Tauchen (1995)].

Assumption 2.2 *The vectors $(z_t, v_t)'$, $t \in \mathbb{N}$ are i.i.d. according to a $N(0, I_2)$ distribution.*

Assumption 2.3 *The process $s_t = (y_t, w_t)'$ is strictly stationary.*

The process is Markovian of order $L_s = max(L_y, L_w)$ with conditional density $p_s(s_t | s_{t-L_s}, \ldots, s_{t-1}, \rho)$ given by the stochastic volatility model,where

$$\rho = (\mu_y, c_1, \ldots, c_{L_y}, r_y, \mu_w, a_1, \ldots, a_{L_w}, r_w)' \tag{2.3}$$

is a vector which contains the free parameters of the stochastic volatility model. The process $\{y_t\}$ is observed whereas $\{w_t\}$ is regarded as latent. Write $p_{y,J}(y_{t-J}, \ldots, y_t|\rho)$ for the implied joint density under the model of a stretch $y_{t-J}, \ldots, y_t$. No general closed-form expressions are available for the moments of y_t, but they can be approximated by Monte Carlo integration.

3. Method-of-moments estimation of an AR(1)-SV model

In this section, we derive analytic expressions for the moments and the estimator of $\theta = (a_1, r_y, r_w)'$ as well as its distributional properties. Let us consider in a first step a simplified version of model (2.1)-(2.2) with $\mu_y = \mu_w = 0$ and $c_j = a_j = 0$, $\forall j \geq 2$. We then have:

$$y_t = c_1 y_{t-1} + \exp(w_t/2) r_y z_t \ , \quad |c_1| < 1 \tag{3.4}$$

$$w_t = a_1 w_{t-1} + r_w v_t \ , \quad |a_1| < 1 \ . \tag{3.5}$$

We shall call the model represented by equations (3.4)-(3.5) the stochastic volatility model with an autoregressive mean part of order one [AR(1)-SV for short]. This specification of the stochastic volatility model comes from Gallant, Hsieh and Tauchen (1995). Let us first introduce some useful notation:

$$u_t(c_1) \equiv y_t - c_1 y_{t-1} \tag{3.6}$$

and

$$v_t(\theta) \equiv \exp(\frac{a_1 w_{t-1} + r_w v_t}{2}) r_y z_t, \ \forall \ t \ . \tag{3.7}$$

with

$$v_t(\theta) = u_t(c_1), \ \forall \ t \ . \tag{3.8}$$

For simplicity of notation, let us call $u_t \equiv u_t(c_1) = v_t(\theta)$. To estimate this AR(1)-SV model above, we consider a two-step method whose first step consists in applying ordinary least squares (OLS) to the mean equation which yields a consistent estimate of the parameter c_1 denoted by $\hat{c}_T$ and the adjusted residuals $\hat{u}_t \equiv u_t(\hat{c}_T)$. Then, we apply in a second step a method of moments to the residuals $\hat{u}_t$ to get the estimate of the parameter $\theta = (a_1, r_y, r_w)'$ of the mean and volatility equations. Jacquier, Polson and Rossi (1994) have derived the expressions of the moments of u_t for particular values of k, namely $E(u_t^2)$, $E(u_t^4)$, $E(u_t^6)$ and $E(u_t^2 u_{t+m}^2)$. We derive them below in the general case for any values of k.

Proposition 3.1 MOMENTS OF THE VOLATILITY PROCESS.

Under Assumptions **2.1**,**2.2**,**2.3**, *with* $\mu_y = \mu_w = 0$ *and* $c_j = a_j = 0$, $\forall\ j \geq 2$. *Then* u_t *has the following moments for even values of* k *and* l:

$$\mu_k(\theta) \equiv E(u_t^k) = r_y^k \frac{k!}{2^{(k/2)}(k/2)!} \exp[\frac{k^2}{8} r_w^2/(1-a_1^2)] \;, \tag{3.9}$$

$$\begin{aligned} \mu_{k,l}(m|\theta) &\equiv E(u_t^k u_{t+m}^l) \\ &= r_y^{k+l} \frac{k!}{2^{(k/2)}(k/2)!} \frac{l!}{2^{(l/2)}(l/2)!} \exp[\frac{r_w^2}{8(1-a_1^2)}(k^2 + l^2 + 2kla_1^m)] \end{aligned} \tag{3.10}$$

The odd moments are equal to zero.

The proofs of the propositions are gathered in the Appendix. In particular, for $k = 2$, $k = 4$ and $k = l = 2$ and $m = 1$, we get as in Jacquier, Polson and Rossi (1994):

$$\mu_2(\theta) = E(u_t^2) = r_y^2 \exp[r_w^2/2(1-a_1^2)] \;, \tag{3.11}$$

$$\mu_4(\theta) = E(u_t^4) = 3r_y^4 \exp[2r_w^2/(1-a_1^2)] \;, \tag{3.12}$$

and

$$\mu_{2,2}(1|\theta) = E[u_t^2 u_{t-1}^2] = r_y^4 \exp[r_w^2/(1-a_1)] \;. \tag{3.13}$$

Solving the above moment equations corresponding to $k = 2$, $k = 4$ and $m = 1$ yields the following proposition.

Proposition 3.2 Method-of-moments estimator.
Under the assumptions of Proposition **3.1**, *we have:*

$$a_1 = \frac{[\log(\mu_{2,2}(1|\theta)) - \log(3) - 4\log(\mu_2) + \log(\mu_4)]}{\log(\frac{\mu_4}{3(\mu_2)^2})} - 1 \;, \tag{3.14}$$

$$r_y = \frac{3^{1/4}\mu_2}{\mu_4^{1/4}} \;, \tag{3.15}$$

$$r_w = \left(\log(\frac{\mu_4}{3(\mu_2)^2})(1-a_1^2)\right)^{1/2} \;. \tag{3.16}$$

Given the latter proposition, it is easy to compute a method-of-moments estimator for $\theta = (a_1, r_y, r_w)'$ replacing the theoretical moments by sample counterparts based on the estimated residuals $\hat{u}_t$. Let $\hat{\theta}_T$ denote the method-of-moments estimator of θ.

Typically, $E(u_t^2)$, $E(u_t^4)$ and $E(u_t^2u_{t-1}^2)$ are approximated by:

$$\hat{\mu}_2 = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 \quad \hat{\mu}_4 = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^4 \ , \quad \hat{\mu}_2(1) = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2\hat{u}_{t-1}^2$$

respectively. Additionally, let $\bar{g}_T(\hat{U}) = \frac{1}{T}\sum_{t=1}^{T} g_t(\hat{U})$ with $g_t(\hat{U}) = (\hat{u}_t^2, \hat{u}_t^4, \hat{u}_t^2\hat{u}_{t-1}^2)'$. Then

$$\bar{g}_T(\hat{U}) = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 \\ \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^4 \\ \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2\hat{u}_{t-1}^2 \end{pmatrix} , \tag{3.17}$$

$\bar{g}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} g_t(\theta)$ with $g_t(\theta) = (v_t^2(\theta), v_t^4(\theta), v_t^2(\theta)v_{t-1}^2(\theta))'$, i.e.

$$\bar{g}_T(\theta) = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T}v_t^2(\theta) \\ \frac{1}{T}\sum_{t=1}^{T}v_t^4(\theta) \\ \frac{1}{T}\sum_{t=1}^{T}v_t^2(\theta)v_{t-1}^2(\theta) \end{pmatrix} , \tag{3.18}$$

with $\mu(\theta) = (\mu_2(\theta), \mu_4(\theta), \mu_{2,2}(1|\theta))'$. In the lemmas below we state some convergence results which will be useful to prove Proposition **3.5**.

Lemma 3.3 *Under the assumptions of Proposition* **3.1**,

$$p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}y_t^2 = \mu_{Y2} \ , \tag{3.19}$$

where $\mu_{Y2} \equiv E(y_t^2) = \frac{\mu_2(\theta)}{1-c_1^2}$, *and we have:*

$$\sqrt{T}(\hat{c}_T - c_1) \xrightarrow{D} N(0, 1-c_1^2) \ . \tag{3.20}$$

Lemma 3.4 *Under the assumptions of Proposition* **3.1**, *we have:*

$$p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}u_t(c_1)y_{t-1} = E[u_t(c_1)y_{t-1}] = 0 \ , \tag{3.21}$$

$$p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}u_t(c_1)^3y_{t-1} = E[u_t(c_1)^3y_{t-1}] = 0 \ , \tag{3.22}$$

$$p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}u_t^2(c_1)u_{t-1}(c_1)y_{t-2} = E[u_t^2(c_1)u_{t-1}(c_1)y_{t-2}] = 0 \ , \tag{3.23}$$

$$p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}u_t(c_1)u_{t-1}^2(c_1)y_{t-1} = E[u_t(c_1)u_{t-1}^2(c_1)y_{t-1}] = 0\ . \tag{3.24}$$

We can now prove the following proposition.

Proposition 3.5 Asymptotic equivalence for $\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta))$.
Under the assumptions of Proposition **3.1**, *the process* $\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta))$ *is asymptotically equivalent to* $\sqrt{T}(\bar{g}_T(\theta) - \mu(\theta))$.

The latter proposition will be useful in deriving the asymptotic distribution of the method-of-moments estimator of θ, but before deriving its asymptotic distribution we need to state the following lemma and proposition.

Lemma 3.6 Expression of covariances.
Let $X_t = (X_{1t}, X_{2t}, X_{3t})'$ *where* $X_{1t} = v_t^2(\theta) - \mu_2(\theta)$,$X_{2t} = v_t^4(\theta) - \mu_4(\theta)$,$X_{3t} = v_t^2(\theta)v_{t-1}^2(\theta) - \mu_{2,2}(1|\theta)$ *and* $\gamma_i(\tau) = Cov(X_{i,t}, X_{i,t+\tau})$, $i = 1, 2, 3$ *then*

$$\gamma_1(\tau) = \mu_2^2(\theta)[\exp(\gamma a_1^\tau) - 1] \tag{3.25}$$
$$\gamma_2(\tau) = \mu_4^2(\theta)[\exp(4\gamma a_1^\tau) - 1] \ \ \forall\tau \geq 1 \tag{3.26}$$
$$\gamma_3(\tau) = \mu_{2,2}^2(1|\theta)[\exp(\gamma(1 + a_1)^2 a_1^{\tau-1}) - 1] \ \ \forall\tau \geq 2 \tag{3.27}$$
$$\tag{3.28}$$

and

$$Cov(w_t, w_{t+\tau}) = a_1^\tau\gamma\ , \tag{3.29}$$

where $\gamma = \frac{r_w^2}{1-a_1^2}$.

Proposition 3.7 Asymptotic distribution of $\sqrt{T}(\bar{g}_T(\theta) - \mu(\theta))$.
Under the assumptions of Proposition **3.1** *and under the assumption that the* 3×3 *matrix* $E[(\sqrt{T}(\bar{g}_T(\theta) - \mu(\theta)))^2]$ *is of full rank for all T, the process* $\sqrt{T}(\bar{g}_T(\theta) - \mu(\theta))$ *is asymptotically distributed as a* $N(0, \Omega^*)$ *variable where* Ω^* *is a positive definite matrix such that* $\Omega^* = \lim_{T\to\infty} E[(\sqrt{T}(\bar{g}_T(\theta) - \mu(\theta)))^2]$.

Next proposition states the asymptotic normality of the method-of-moments estimator of $\theta = (a_1, r_y, r_w)'$ a subvector of $\rho' = (c, \theta')$.

Proposition 3.8 Asymptotic distribution of the method-of-moments estimator.
Under the assumptions of Proposition **3.1**, *the method-of-moments estimator* $\hat{\theta}_T(\Omega)$ *is such that:*

$$\sqrt{T}[\hat{\theta}_T(\Omega) - \theta] \xrightarrow{D} N(0, W(\Omega)) \tag{3.30}$$

where

$$W(\Omega) = \left(\frac{\partial\mu'}{\partial\theta}(\theta)\Omega\frac{\partial\mu}{\partial\theta'}(\theta)\right)^{-1}\frac{\partial\mu'}{\partial\theta}(\theta)\Omega\Omega^*\Omega\frac{\partial\mu}{\partial\theta'}(\theta)$$
$$\left(\frac{\partial\mu'}{\partial\theta}(\theta)\Omega\frac{\partial\mu}{\partial\theta'}(\theta)\right)^{-1} \tag{3.31}$$

As usual, there is an optimal choice of this matrix, i.e. a choice which minimizes $W(\Omega)$.

Proposition 3.9 OPTIMAL WEIGHTING MATRIX.
Under the assumptions of Proposition **3.1**, *the optimal choice of the* Ω *matrix is:* $\Omega = \Omega^{*-1}$ *and*

$$W^* = W(\Omega^{*-1}) = \left(\frac{\partial\mu'}{\partial\theta}(\theta)\Omega^{*-1}\frac{\partial\mu}{\partial\theta'}(\theta)\right)^{-1}. \tag{3.32}$$

The optimal estimator thus obtained is denoted by $\hat{\theta}_T$. When the dimensions of μ and θ are the same, we have $W(\Omega) = W^*$, $\forall\ \Omega$ and

$$W^* = \left(\frac{\partial\mu'}{\partial\theta}(\theta)\right)^{-1}\Omega^*\left(\frac{\partial\mu}{\partial\theta'}(\theta)\right)^{-1}.$$

It is the asymptotic variance-covariance matrix of the estimator solution of

$$\bar{g}_T(\hat{U}) = \mu(\theta)\ .$$

A consistent estimator of W^* is obtained as soon as we have a consistent estimator of Ω^*. A consistent estimator of Ω^* can be easily obtained [see Newey and West (1987)] by a fixed-bandwith Bartlett kernel estimator, i.e.:

$$\hat{\Omega}^*(\theta) = \Gamma_0 + \sum_{k=1}^{K}(1 - \frac{k}{K+1})(\Gamma_k + \Gamma_k') \tag{3.33}$$

with

$$\Gamma_k = \frac{1}{T}\sum_{t=k+1}^{T}[g_{t-k}(\theta) - \bar{g}_T(\theta)][g_t(\theta) - \bar{g}_T(\theta)]' \tag{3.34}$$

with θ replaced by any consistent estimator $\tilde{\theta}_T$ of θ,

$$\hat{\Omega}^*(\tilde{\theta}_T) = \hat{\Gamma}_0 + \sum_{k=1}^{K}(1 - \frac{k}{K+1})(\hat{\Gamma}_k + \hat{\Gamma}_k') \tag{3.35}$$

with

$$\hat{\Gamma}_k = \frac{1}{T}\sum_{t=k+1}^{T}[g_{t-k}(\tilde{\theta}_T) - \bar{g}_T(\tilde{\theta}_T)][g_t(\tilde{\theta}_T) - \bar{g}_T(\tilde{\theta}_T)]' \tag{3.36}$$

since

$$\bar{g}_T(\tilde{\theta}_T) = \frac{1}{T}\sum_{t=1}^{T} g_t(\tilde{\theta}_T) \xrightarrow{p} \mu(\theta_0) = E[g_t(\theta_0)]$$

and

$$\frac{1}{T}\sum_{t=1}^{T} g_{t-k}(\tilde{\theta}_T) g_t(\tilde{\theta}_T)' \xrightarrow{p} E[g_{t-k}(\theta) g_t(\theta)']$$

since the perturbation vectors have been shown to be strictly stationary and ergodic [see proof of Proposition **3.7** in Appendix].

Therefore a consistent estimator of W^* is given by:

$$\hat{W}^* = \left(\frac{\partial \mu'}{\partial \theta}(\hat{\theta}_T)\hat{\Omega}^{*-1}(\tilde{\theta}_T)\frac{\partial \mu}{\partial \theta'}(\hat{\theta}_T)\right)^{-1}. \tag{3.37}$$

4. Hypothesis tests

We assume that the parameter $\theta = (a, r_y, r_w)'$ is partitioned into

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$

with $\theta_1 \overset{def}{=} a$. We consider the null hypothesis $H_0 : \ a = 0$ which corresponds to test the absence of long memory in the volatility. To define these tests we have to introduce the optimal unconstrained estimator:

$$\begin{pmatrix} \hat{\theta}_{1T} \\ \hat{\theta}_{2T} \end{pmatrix} = \hat{\theta}_T$$

and the optimal constrained estimator obtained by optimizing the criterion submitted to $\theta_1 = 0$. This estimator is denoted by:

$$\begin{pmatrix} 0 \\ \hat{\theta}^c_{2T} \end{pmatrix} = \hat{\theta}^c_T \ .$$

The Wald statistic is defined as

$$\xi^W_T = T(\hat{\theta}_{1T})'\hat{W}^{*-1}_1(\hat{\theta}_{1T}) \tag{4.38}$$

where $\hat{W}^*_1$ is a consistent estimator of the asymptotic covariance-variance matrix of $\sqrt{T}\hat{\theta}_{1T}$. W^*_1 is defined at equation (A.128).

The score statistic is defined from the gradient of the objective function with respect to θ_1 evaluated at the constrained estimator. This gradient is given by:

$$\mathcal{D}_T = \frac{\partial \mu'}{\partial \theta_1}(\hat{\theta}^c_T)\Omega^{*-1}(\mu(\hat{\theta}^c_T) - \bar{g}_T(\hat{U})) \tag{4.39}$$

and the test statistic is

$$\xi^S_T = T\mathcal{D}'_T\mathcal{A}\mathcal{D}_T \ . \tag{4.40}$$

where $\mathcal{A}$ is a consistent estimator of the inverse of the covariance matrix of $\sqrt{T}\mathcal{D}_T$ whose covariance matrix is defined at equation (A.130). Finally, we can introduce the difference between the optimal values of the objective function that we will call the LR-type test in the simulations:

$$\xi^C_T = T[M^*_T(\hat{\theta}^c_T) - M^*_T(\hat{\theta}_T)] \tag{4.41}$$

where the criterion to be minimized is:

$$M^*_T(\theta) \stackrel{def}{=} [\bar{g}_T(\hat{U}) - \mu(\theta)]'\hat{\Omega}^{*-1}[\bar{g}_T(\hat{U}) - \mu(\theta)] \tag{4.42}$$

Proposition 4.1 Asymptotic distribution of the three classic tests.

Under the assumptions of Proposition **3.1**, *the test statistics* ξ^W_T, ξ^S_T, *and* ξ^C_T *are asymptotically equivalent under the null hypothesis, and have the common distribution* $\chi^2(1)$.

We also consider the $C(\alpha)$-type test statistic defined by:

$$PC(\tilde{\theta}^c_T) = T[\mu(\tilde{\theta}^c_T) - \bar{g}_T(\hat{U})]'\tilde{W}_0[\mu(\tilde{\theta}^c_T) - \bar{g}_T(\hat{U})] \tag{4.43}$$

where

$$\begin{aligned}\tilde{W}_0 \;=\; & \tilde{I}_0^{-1}\tilde{J}_0(\tilde{J}_0'\tilde{I}_0^{-1}\tilde{J}_0)^{-1} \\ & [\tilde{P}_0(\tilde{J}_0'\tilde{I}_0^{-1}\tilde{J}_0)^{-1}\tilde{P}_0']^{-1} \\ & \tilde{P}_0(\tilde{J}_0'\tilde{I}_0^{-1}\tilde{J}_0)^{-1}\tilde{J}_0'\tilde{I}_0^{-1} \,, \end{aligned} \tag{4.44}$$

with

$$\tilde{J}_0 = J(\tilde{\theta}_T^c) = \frac{\partial \mu}{\partial \theta'}(\tilde{\theta}_T^c) \,, \tag{4.45}$$

$$\tilde{I}_0^{-1} = I(\tilde{\theta}_T^c)^{-1} = \Omega^*(\tilde{\theta}_T^c)^{-1} \tag{4.46}$$

and $\tilde{P}_0 = P(\tilde{\theta}_T^c)$ where $P(\theta)$ corresponds to the derivative of the constraints w.r.t. the parameters of interest θ evaluated at any root-n consistent estimator of θ that satisfies the constraint $a = 0$. For our concern, $\tilde{\theta}_T^c$ will be obtained by setting $a = 0$ in the analytical expressions of the unrestricted method-of-moments estimator $\hat{\theta}_T$ given at equations (3.14) to (3.16). It is known [see Dufour and Trognon (2001, p.8, Proposition 3.1)] that the $C(\alpha)$-type test statistic is asymptotically distributed as a $\chi^2(1)$ variable under the null hypothesis.

5. Monte Carlo testing

The technique of Monte Carlo tests has originally been suggested by Dwass (1957) for implementing permutation tests, and did not involve nuisance parameters. This technique has been later extended by Barnard (1963) and Birnbaum (1974). This technique has the great attraction of providing *exact* (randomized) tests based on any statistic whose finite sample distribution may be intractable but can be simulated.

We study here the case where the distribution of the test statistic S depends on nuisance parameters. For the test statistics exposed in section 4, their asymptotic distribution is asymptotically pivotal (Chi-square distribution), but their finite sample distribution remains unknown. At this stage, we need to make an effort of formalization to clearly expose the procedure. We consider a family of probability spaces $\{(\mathcal{Z}, \mathcal{A}_{\mathcal{Z}}, P_\rho) : \rho \in \Omega\}$ and suppose that S is a real valued $\mathcal{A}_{\mathcal{Z}}$-measurable function whose distribution is determined by $P_{\bar{\rho}}$ where $\bar{\rho}$ is the "true" parameter vector. We wish to test the hypothesis

$$H_0 : \bar{\rho} \in \Omega_0,$$

where Ω_0 is a nonempty subset of Ω. We take a critical region of the form $S \geq c$, where c is a constant which does not depend on ρ. The critical region $S \geq c$ has *level* α if and only if

$$P_\rho[S \geq c] \leq \alpha, \forall \rho \in \Omega_0,$$

or equivalently,

$$\sup_{\rho \in \Omega_0} P_\rho[S \geq c] \leq \alpha.$$

Firthermore, $S \geq c$ has *size* α when

$$\sup_{\rho \in \Omega_0} P_\rho[S \geq c] = \alpha.$$

If we define the distribution and p-value functions,

$$F[x|\rho] = P_\rho[S \leq x], x \in \bar{R},$$

$$G[x|\rho] = P_\rho[S \geq x], x \in \bar{R},$$

where $\rho \in \Omega$, it is easy to see that the critical regions

$$\sup_{\rho \in \Omega_0} G[S|\rho] \leq \alpha(c),$$

where $\alpha(c) \equiv \sup_{\rho\in\Omega_0} G[c|\rho]$, and

$$S \geq \sup_{\rho\in\Omega_0} F^{-1}[(1 - G[c|\rho])^{+}|\rho] \equiv \bar{c}$$

are equivalent to $S \geq c$ in the sense that $c \leq \bar{c}$.

We consider a real random variable S_0 and random vectors of the form

$$S(N, \rho) = (S_1(\rho), \ldots, S_N(\rho))', \rho \in \Omega,$$

all defined on a common probability space $(\mathcal{Z}, \mathcal{A}_{\mathcal{Z}}, P)$, such that the variables $S_0, S_1(\bar{\rho}), \ldots, S_N(\bar{\rho})$ are i.i.d. or exchangeable for some $\bar{\rho} \in \Omega$, each one with distribution function $F[x|\bar{\rho}] = P[S_0 \leq x]$. Typically, S_0 will refer to a test statistic computed from the observed data when the true parameter vector is $\bar{\rho}$ (i.e., $\rho = \bar{\rho}$), while $S_1(\rho), \ldots, S_N(\rho)$ will refer to i.i.d replications of the test statistic obtained independently (e.g., by simulation) under the assumption that the parameter vector is ρ (i.e., $P[S_i(\rho) \leq x] = F[x|\rho]$). In other words, the observed statistic S_0 is simulated by first generating an "observation" vector y according to

$$y = g(\rho, z, v) \tag{5.47}$$

where the function g is bivariate for our AR(1)-SV model, and corresponds to equations (3.4) and (3.5), with $\rho = (c, \theta)'$, $\theta = (a, r_y, r_w)'$. The perturbations z and v have known distributions, which can be simulated ($N(0, 1)$ or student, or mixtures, e.g.) and then computing

$$S(\rho) \equiv S[g(\rho, z, v)] \equiv g_S(\rho, z, v) \ . \tag{5.48}$$

The observed statistic S_0 is then computed as $S_0 = S[g(\bar{\rho}, z_0, v_0)]$ and the simulated statistics $S_i(\rho) = S[g(\rho, z_i, v_i)]$, $i = 1, \ldots, N$ where the random vectors $z_0, z_1, \ldots, z_N$ are i.i.d. (or exchangeable) and $v_0, v_1, \ldots, v_N$ are i.i.d. (or exchangeable) as well.

The technique of Monte Carlo tests provides a simple method allowing one to replace the theoretical distribution $F(x|\rho)$ by its sample analogue based on $S_1(\rho), \ldots, S_N(\rho)$:

$$\hat{F}_N[x; S(N, \rho)] = \frac{1}{N}\sum_{i=1}^{N} s(x - S_i(\rho)) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{[0,\infty]}(x - S_i(\rho))$$

where $s(x) = 1_{[0,\infty]}(x)$ and $1_A(x)$ is the indicator function associated with the set A.

We also consider the corresponding sample tail area function:

$$\hat{G}_N[x; S(N,\rho)] = \frac{1}{N}\sum_{i=1}^{N} s(S_i(\rho) - x)\cdot$$

and the p-value function

$$\hat{p}_N[x|\rho] = \frac{N\hat{G}_N[x|\rho] + 1}{N+1} \ .$$

The sample distribution function is related to the ranks $R_1, \cdots, R_N$ of the variables $S_1(\rho), \ldots, S_N(\rho)$ (when put in ascending order) by the expression:

$$R_j = N\hat{F}_N[S_j; S(N,\rho)] = \sum_{i=1}^{N} s(S_j(\rho) - S_i(\rho)), \ j = 1, \ldots, N.$$

The central property which is exploited here is the following: to obtain critical values or compute p-values, the "theoretical" null distribution $F[x|\bar{\rho}]$ can be replaced by its simulation-based "estimate" $\hat{F}_N[x|\rho] \equiv \hat{F}_N[x; S(N,\rho)]$ in a way that will preserve the level of the test in *finite samples, irrespective of the number N of replications used.*

Thus, in this framework, Dufour (1995) states the finite sample validity of Monte Carlo tests when the p-value function is maximized over the entire set of the nuisance parameters as it is formulated in the reported proposition below [see Dufour (1995, p.13, Proposition 4.1].

Proposition 5.1 Validity of MMC tests when ties have zero probability.

Under the above assumptions and notations, set
$S_0(\bar{\rho}) = S_0$ and suppose that

$$P[S_i(\bar{\rho}) = S_j(\bar{\rho})] = 0, \quad \textit{for } i \neq j, \ i, j = 0, 1, \ldots, N.$$

If $\bar{\rho} \in \Omega_0$, then for $0 \leq \alpha_1 \leq 1$,

$$\begin{aligned} P[\sup\{\hat{G}_N[S_0|\rho] : \rho \in \Omega_0\} \leq \alpha_1] &\leq P[\inf\{\hat{F}_N[S_0|\rho] : \rho \in \Omega_0\} \geq 1 - \alpha_1] \\ &\leq \frac{I[\alpha_1 N] + 1}{N+1} \end{aligned}$$

where

$$P[\inf\{\hat{F}_N[S_0|\rho] : \rho \in \Omega_0\} \geq 1 - \alpha_1] = P[S_0 \geq \sup\{\hat{F}_N^{-1}[1 - \alpha_1|\rho] : \rho \in \Omega_0\}]$$

for $0 < \alpha_1 < 1$, *and*

$$P[\sup\{\hat{p}_N[S_0|\rho] : \rho \in \Omega_0\} \leq \alpha] \leq \frac{I[\alpha(N+1)]}{N+1}, for\ 0 \leq \alpha \leq 1.$$

Following the latter proposition, if we choose α_1 and N so that

$$\alpha = \frac{I[\alpha_1 N] + 1}{N+1} \tag{5.49}$$

is the desired significance level, the critical region $\sup\{\hat{G}_N[S_0|\rho] : \rho \in \Omega_0\} \leq \alpha_1$ has level α irrespective of the presence of nuisance parameters in the distribution of the test statistic S under the null hypothesis $H_0 : \bar{\rho} \in \Omega_0$. The same also holds if we use the (almost) equivalent critical regions $\inf\{\hat{F}_N[S_0|\rho] : \rho \in \Omega_0\} \geq 1 - \alpha_1$ or $S_0 \geq \sup\{\hat{F}_N^{-1}[1 - \alpha_1|\rho] : \rho \in \Omega_0\}$. Dufour call such tests maximized Monte Carlo (MMC) tests. The function $\hat{G}_N[S_0|\rho]$ (or $\hat{p}_N[S_0|\rho]$) is then maximized with respect to $\rho \in \Omega_0$, keeping the observed statistic S_0 and the simulated disturbance vectors $z_1, ..., z_N$ and $v_1, ..., v_N$ fixed. The function $\hat{G}_N[S_0|\rho]$ is a step-type function which typically has zero derivatives almost everywhere, except on isolated points (or manifolds) where it is not differentiable. So it cannot be maximized with usual derivative-based algorithms. However, the required maximizations can be performed by using appropriate optimization algorithms that do not require differentiability, such as *simulated annealing*. For further discussion of such algorithms, the reader may consult Goffe, Ferrier, and Rogers(1994).

On the other hand, Dufour (1995) also proposes simplified (asymptotically justified) approximate versions of the Monte Carlo tests where this time the p-value function is evaluated at a consistent point estimate, which defines a Bootstrap version, or a consistent set estimate of ρ, which defines asymptotic Monte Carlo tests based on consistent set estimators. The author shows [see Dufour, (1995, p.16, Proposition 5.1 and p.19, Proposition 6.3)] that both tests are asymptotically valid in the sense that they have the correct level α asymptotically and the estimated p-values converge to the true p-values. He also assesses the validity of the MMC tests and the asymptotic Monte Carlo tests based on consistent set estimators for general distributions , when ties have non-zero probability [see Dufour, (1995, p.14, Proposition 4.2 and p.17, Proposition 5.2)].

It is this technique of Monte Carlo tests in their maximized and Bootstrap versions which will be applied in section 6 to compare their level and power with those of the standard asymptotic tests of section 4.

6. Implementation results

Here we test the null hypothesis of no-persistence in the volatility, which corresponds to $H_0 : a = 0$ against the alternative $H_1 : a = 0.9$. The nominal level of the tests has been set to $\alpha = 5\%$. M represents the number of replications to evaluate the frequence of rejection of both hypotheses, and N represents the number of simulated statistics used in the Monte Carlo tests. T is the sample size of the serie y_t whose data generating process is assumed to be specified as in equations (3.4)-(3.5). Implementation is performed with the GAUSS software.

The Wald statistic as defined at equation (4.38) is evaluated at the unrestricted method-of-moments estimator $\hat{\theta}_{1T}$. The Score statistic as defined at equation (4.40) is evaluated at the restricted estimator $\hat{\theta}_T^c$ which minimizes the criterion $M_T^*(\theta)$ defined at equation (4.42) submitted to the constraint $a = 0$ whereas $\tilde{\theta}_T^c$ represents another restricted estimator of θ obtained by setting $a = 0$ in the analytical expressions of the unrestricted method-of-moments estimator $\hat{\theta}_T$ given at equations (3.14) to (3.16). The $C(\alpha)$-type statistic as defined at equation (4.43) is evaluated at this restricted estimator $\tilde{\theta}_T^c$ of θ. Additionally, the LR-type test statistic corresponds to the difference between the optimal values of the objective function. Let $LR(\hat{\Omega}) \equiv \xi_T^C$ [see equation (4.41)] where $\hat{\Omega} \equiv \Omega(\hat{\theta}_T)$. The weighting matrix $\hat{\Omega}$ is estimated by a kernel estimator with a fixed-Bandwith Bartlett Kernel, where the lag truncation parameter K has been set to $K = 5$ as defined through equations (3.35) and (3.36).

Let S denote the test statistic which alternatively will take the form of one of the four test statistics earlier mentionned and let S_0 denote the statistic computed from the observed data (Standard and Poor's Composite Price Index) or the "pseudo-true" data obtained by simulation under the hypothesis to be tested. The critical regions used to perform the tests are of the form:

$$\mathcal{R}_c = \{S_0 > \chi^2_{1-\alpha}(1) = 3.84\}$$

for the standard asymptotic tests, and of the form:

$$\mathcal{R}_c = \{\hat{p}_N[S_0|\hat{\rho}_T^c] \leq \alpha\}$$

with the p-value function

$$\hat{p}_N[S_0|\rho] = \frac{N\hat{G}_N[S_0|\rho] + 1}{N + 1} \, ,$$

and the tail area function

$$\hat{G}_N[S_0; S(N, \rho)] = \frac{1}{N} \sum_{i=1}^{N} s(S_i(\rho) - S_0),$$

for the Bootstrap tests where the p-value function may be evaluated at any consistent restricted estimator of $\rho = (c, \theta')' = (c, a, r_y, r_w)'$. The simulated statistics $S_i(\rho)$ $i = 1, ..., N$ will always be evaluated under the null hypothesis in the Monte Carlo tests whatever the hypothesis to be tested. α has been set to $\alpha = 5\%$. The Bootstrap version of the Monte Carlo tests whose p-value function is evaluated at a consistent point estimate of the nuisance parameters follows the methodology explicited in section 5.

6.1. Test level

Here we study the empirical frequence of rejection of the null hypothesis $H_0 : a = 0$ and compare it to the nominal level fixed at $\alpha = 5\%$.

LEVELS in % (under H_0)						
	Asymptotic tests					
	T=50	T=100	T=200	T=500	T=1000	T=2000
Wald	0.4	1.3	1.8	4.4	2.4	3.2
$Score(\hat{\Omega}_C)$	14.3	7.6	5.4	4.2	3.2	3
$LR(\hat{\Omega})$	25.8	17.9	13.7	7.4	3.9	3.7
$C(\alpha)$	3.7	2.6	2.9	3	2.9	2.9

LEVELS in % (under H_0)						
	Bootstrap tests					
	T=50	T=100	T=200	T=500	T=1000	T=2000
Wald	5.1	5.1	4.3	2	0.8	3.1
$Score(\hat{\Omega}_C)$	3.1	1.7	2	4.3	4.4	2.9
$LR(\hat{\Omega}_{NC})$	3.5	3.2	2.6	1.4	0.5	2.9
$C(\alpha)$	4.7	4.4	4.9	6.3	5.4	4

Note that the Bootstrap test which is a simplified (asymptotically justified) version of the Maximized Monte Carlo test, reduces drastically the size distorsions observed for its standard asymptotic counterpart, mostly for the score test statistic. The $C(\alpha)$ test performs quite well and is attractive in this context since it does not require any computer optimization to get the restricted estimate of θ. However the level of the $LR(\hat{\Omega})$ test remains over 5% for a sample size of $T = 1000$, which suggests a maximized version of the Monte Carlo test to control the level in this case.

6.2. Test power

Here we study the power of the tests, that is the empirical frequence of rejection of the null hypothesis $H_0 : a = 0$ when the data have been generated under alternative hypothesis $H_1 : a = 0.8$. The first table below reports the power of the standard asymptotic tests whose size has been corrected with respect to the corresponding simulated critical values which yield exact 5%-level tests under the null hypothesis.

Simulated critical values						
	MM=10000 replications					
	T=50	T=100	T=200	T=500	T=1000	T=2000
Wald	0.9619	1.4615	2.1764	3.5563	3.0587	3.0344
$Score(\hat{\Omega}_C)$	8.8501	5.5501	3.8858	3.3685	3.0459	2.9812
$LR(\hat{\Omega})$	39.6310	19.0368	11.7330	5.0693	3.4685	3.0480
$C(\alpha)$	3.1991	3.0226	2.8863	2.9096	2.8879	2.9133

POWER in % (under H_1)						
	Size-corrected Asymptotic tests					
	T=50	T=100	T=200	T=500	T=1000	T=2000
Wald	12.6	17.1	28.8	50.2	82.4	92.5
$Score(\hat{\Omega}_C)$	16.9	21.3	42	78.4	95.8	99.6
$LR(\hat{\Omega})$	13	8.2	13.4	55.8	87.3	96.6
$C(\alpha)$	17.2	30.1	50.2	81.5	96	99.5

POWER in % (under H_1)						
	Bootstrap tests ($N = 99$)					
	T=50	T=100	T=200	T=500	T=1000	T=2000
Wald	10.9	12.5	19.9	43.2	68.1	83.5
$Score(\hat{\Omega}_C)$	12.5	10.8	19.5	49.8	76.2	88.7
$LR(\hat{\Omega})$	10.6	6.8	12.8	53.7	81.9	91.4
$C(\alpha)$	15.9	27.1	42.6	73.9	93.5	98.5

We do not prescribe these methods when the sample size is very small (e.g. $T = 50$), the tests do have ver little power and in this case a maximized Monte Carlo test could improve the results. Both test procedures have more power when the sample size grows which is intuitive since both tests are asymptotically justified. In the standard asymptotic procedure, the likelihood ratio type (LR) and $C(\alpha)$ tests are the most powerful when the sample size is large. Once again, the $C(\alpha)$ test is the most powerful one in the Bootstrap procedure where it reaches and exceeds a power of 90% while being the easiest to implement since it does not require in our case any optimization procedure.

7. Empirical application

In this subsection we test the null hypothesis of no-persistence in the volatility from real data (Standard and Poor's Composite Price Index (SP), 1928-87).

7.1. Data

The data have been provided by Tauchen where Efficient Method of Moments have been used by Gallant, Hsieh and Tauchen to fit a standard stochastic volatility model and various extensions. The data to which we fit the univariate stochastic volatility model is a long time series comprised of 16,127 daily observations, $\{\tilde{y}_t\}_{t=1}^{16,127}$, on adjusted movements of the Standard and poor's Composite Price Index, 1928-87. The raw series is the Standard and Poor's Composite Price Index (SP),daily, 1928-87. We use a long time series, because, among other things, we want to investigate the long-term properties of stock market volatility through a persistence test. The raw series is converted to a price movements series, $100[\log(SP_t) - \log(SP_{t-1})]$, and then adjusted for systematic calendar effects, that is, systematic shifts in location and scale due to different trading patterns across days of the week, holidays, and year-end tax trading. This yields a variable we shall denote y_t.

7.2. Results

The unrestricted estimated value of ρ from the data is:

$$\hat{\rho}_T = (0.129, 0.926, 0.829, 0.427)'$$

where the method-of-moments estimated value of a corresponds to $\hat{a}_T = 0.926$. We may conjecture that there is some persistence in the data during the period 1928-87 what is statistically checked by performing the tests below. The restricted estimated values of ρ from the data are:

$$\hat{\rho}_T^c = (0.129, 0, 0.785, 1.152)'$$

and

$$\tilde{\rho}_T^c = (0.129, 0, 0.829, 1.133)' .$$

Note the large discrepancy between the unrestricted and restricted estimated values of r_w.

data				
	Asymptotic tests	*Bootstrap tests*		
	S_0	N=19	N=99	N=999
Wald	206.03	0.05	0.01	0.001
$Score(\hat{\Omega}_C)$	1039.04	0.05	0.01	0.001
$LR(\hat{\Omega})$	63.20	0.05	0.01	0.001
$C(\alpha)$	854.55	0.05	0.01	0.001

All standard asymptotic tests reject indeed the null hypothesis of no-persistence in the volatility since $S_0 > \chi^2_{1-\alpha}(1) = 3.84$ as well as all the Bootstrap tests whose p-value is equal or less than 5%, whatever length of the simulated statistics is used to implement them.

8. Concluding remarks

The $C(\alpha)$ test outperforms the other types of tests while being the easiest to implement since it does not require in our framework any optimization procedure. It has good statistical properties: a good level and a high power for sufficiently large sample sizes. On the other hand, the Monte Carlo tests in general appear as a good alternative to the standard asymptotic tests, specifically when the standard asymptotic approach fails - unit root specification or small-sample tests where the distribution of the test statistic in unknown. We may consider as further reasearch an extension of our approach to asymetric distributions such as the asymetric student distribution and a willingness to test the hypothesis of leverage effect in the stochastic volatility model.

A. Appendix

Proof of Proposition 3.1
First, we recall that if $U \sim N(0,1)$ then $E(U^{2p+1}) = 0$, $\forall\ p \in \mathbb{N}$ and $E(U^{2p}) = (2p)!/[2^p p!]$ $\forall\ p \in \mathbb{N}$ [see Gouriéroux, Monfort, p.518, vol.2]. Under Assumptions **2.1**,**2.2**,**2.3** with a stationary AR(1) specification for w_t, using the definition of u_t, we get:

$$\begin{aligned}
E(u_t^k) &= r_y^k E(z_t^k) E[\exp(k w_t/2)] \\
&= r_y^k \frac{k!}{2^{(k/2)}(k/2)!} \exp(\frac{k^2}{4} r_w^2/2(1-a_1^2)) \\
&= r_y^k \frac{k!}{2^{(k/2)}(k/2)!} \exp(\frac{k^2}{8} r_w^2/(1-a_1^2)) \qquad \text{(A.50)}
\end{aligned}$$

where the second equality uses the definition of the gaussian Laplace transform of $w_t \sim N(0, \frac{r_w^2}{1-a_1^2})$ and of the moments of the $N(0,1)$ z_t variable. Let us now calculate the cross-product:

$$\begin{aligned}
E[u_t^k u_{t+m}^l] &= E[r_y^{k+l} z_t^k z_{t+m}^l \exp(k\frac{w_t}{2} + l\frac{w_{t+m}}{2})] \\
&= r_y^{k+l} E(z_t^k) E(z_{t+m}^l) E[\exp(k\frac{w_t}{2} + l\frac{w_{t+m}}{2})] \\
&= r_y^{k+l} \frac{k!}{2^{(k/2)}(k/2)!} \frac{l!}{2^{(l/2)}(l/2)!} \exp\left(\frac{r_w^2}{8(1-a_1^2)}(k^2 + l^2 + 2kla_1^m)\right)
\end{aligned}$$

where $E(w_t) = 0$, $Var(w_t) = \frac{r_w^2}{1-a_1^2}$ and

$$\begin{aligned}
Var(k\frac{w_t}{2} + l\frac{w_{t+m}}{2}) &= \frac{k^2}{4} Var(w_t) + \frac{l^2}{4} Var(w_{t+m}) + 2\frac{k}{2}\frac{l}{2} Cov(w_t, w_{t+m}) \\
&= \frac{r_w^2}{4(1-a_1^2)}(k^2 + l^2 + 2kla_1^m) \ . \qquad \text{(A.51)}
\end{aligned}$$

□

Proof of Proposition 3.2
Taking the ratio of equation (3.12) on equation (3.11) to the square produces

$$\frac{E(u_t^4)}{(E(u_t^2))^2} = 3\exp(r_w^2/(1-a_1^2)) \ , \qquad \text{(A.52)}$$

i.e.

$$Q \equiv (r_w^2/(1-a_1^2)) = \log\left(\frac{E(u_t^4)}{3(E(u_t^2))^2}\right) . \tag{A.53}$$

Inserting $Q \equiv (r_w^2/(1-a_1^2))$ in equation (3.11) yields

$$r_y = \left(\frac{E(u_t^2)}{\exp(Q/2)}\right)^{1/2} = \frac{3^{1/4}E(u_t^2)}{E(u_t^4)^{1/4}} . \tag{A.54}$$

From equation (3.13) we have

$$\exp(\frac{r_w^2}{(1-a_1)}) = \frac{E[u_t^2 u_{t-1}^2]}{r_y^4} \tag{A.55}$$

which, after a few manipulations, yields

$$1 + a_1 = \frac{[\log(E[u_t^2 u_{t-1}^2]) - 4\log(r_y)]}{Q} \tag{A.56}$$

or either

$$a_1 = \frac{[\log(E[u_t^2 u_{t-1}^2]) - \log(3) - 4\log(E[u_t^2]) + \log(E[u_t^4])]}{\log\left(\frac{E[u_t^4]}{3(E[u_t^2])^2}\right)} - 1 . \tag{A.57}$$

From the expressions of $Q \equiv r_w^2/(1-a_1^2)$ at equation (A.53) and that of a_1 above we can deduce

$$r_w = \left(\log\left(\frac{E[u_t^4]}{3(E[u_t^2])^2}\right).(1-a_1^2)\right)^{1/2} . \tag{A.58}$$

$\square$

Proof of Lemma 3.3

To prove the result, we shall show first that the process $\{y_t^2 - \mu_{Y2}\ , t \in \mathbb{N}\}$ is a L_1-mixingale w.r.t. the subfields $\mathcal{F}_t = \sigma(s_t, s_{t-1}, \ldots)$ where $s_t = (y_t, w_t)'$. In a second step, we shall show that the sequence $\{y_t^2 - \mu_{Y2}\ , t \in \mathbb{N}\}$ is uniformly integrable. Finally we shall apply the Law of Large numbers (L.L.N.) for L_1-mixingales to the sequence $\{y_t^2 - \mu_{Y2}\ , t \in \mathbb{N}\}$ to get the desired result.

For the L_1-mixingale property, we note first that:

$$\begin{aligned} y_t^2 &= (c_1 y_{t-1} + \exp(w_t/2) r_y z_t)^2 \\ &= c_1^2 y_{t-1}^2 + \exp(w_t) r_y^2 z_t^2 + 2c_1 y_{t-1} \exp(w_t/2) r_y z_t \end{aligned} \tag{A.59}$$

hence, iterating backwards on y_t^2,

$$
\begin{aligned}
y_t^2 &= (c_1^2)^m y_{t-m}^2 + (c_1^2)^{m-1}[\exp(w_{t-m+1}) r_y^2 z_{t-m+1}^2 + 2c_1 y_{t-m} \exp(\frac{w_{t-m+1}}{2}) r_y z_{t-m+1}] \\
& \quad + \ldots + (c_1^2)^0 [\exp(w_t) r_y^2 z_t^2 + 2c_1 y_{t-1} \exp(w_t/2) r_y z_t] \; . \qquad \text{(A.60)}
\end{aligned}
$$

Besides, by Assumptions **2.2** and **2.3** we have:

$$
\mu_{Y2} \equiv E y_t^2 = \frac{\mu_2}{1 - c_1^2} \; .
$$

Therefore, taking the conditional expectation from both sides of equation (A.60), we get:

$$
\begin{aligned}
E(y_t^2 - \mu_{Y2} | \mathcal{F}_{t-m}) &= (c_1^2)^m y_{t-m}^2 - \mu_{Y2} \\
& \quad + \sum_{j=0}^{m-1} (c_1^2)^j \left\{ E\left[\exp(w_{t-j}) r_y^2 z_{t-j}^2 + 2c_1 y_{t-1-j} \exp(w_{t-j}/2) r_y z_{t-j} | \mathcal{F}_{t-m} \right] \right\} \\
&= (c_1^2)^m y_{t-m}^2 - \mu_{Y2} + \sum_{j=0}^{m-1} (c_1^2)^j \left\{ r_y^2 E(z_{t-j}^2 | \mathcal{F}_{t-m}) \right. \\
& \quad \left. E\left[\exp(a_1^{m-j} w_{t-m} + \sum_{l=0}^{m-j-1} a_1^l r_w v_{t-j-l}) | \mathcal{F}_{t-m} \right] \right\} \\
&= (c_1^2)^m y_{t-m}^2 - \mu_{Y2} + \sum_{j=0}^{m-1} (c_1^2)^j \left\{ r_y^2 E(z_{t-j}^2 | \mathcal{F}_{t-m}) \exp(a_1^{m-j} w_{t-m}) \right. \\
& \quad \left. E\left[\prod_{l=0}^{m-j-1} \exp(a_1^l r_w v_{t-j-l}) | \mathcal{F}_{t-m} \right] \right\} \qquad \text{(A.61)}
\end{aligned}
$$

hence, since $z_t \overset{i.i.d.}{\sim} N(0,1)$ and $v_t \overset{i.i.d.}{\sim} N(0,1)$,

$$
E(y_t^2 - \mu_{Y2} | \mathcal{F}_{t-m}) = (c_1^2)^m y_{t-m}^2 - \mu_{Y2} + \sum_{j=0}^{m-1} (c_1^2)^j \left\{ r_y^2 \exp(a_1^{m-j} w_{t-m}) \prod_{l=0}^{m-j-1} E \exp(a_1^l r_w v_{t-j-l}) \right\}
$$

or equivalently,

$$
E(y_t^2 - \mu_{Y2} | \mathcal{F}_{t-m}) = (c_1^2)^m y_{t-m}^2 - \mu_{Y2} + \sum_{j=0}^{m-1} (c_1^2)^j \left\{ r_y^2 \exp(a_1^{m-j} w_{t-m}) \prod_{l=0}^{m-j-1} \exp(\frac{1}{2} a_1^{2l} r_w^2) \right\} \; .
$$

Therefore, taking the unconditional expectation of the absolute value, we deduce

$$
\begin{aligned}
E|E(y_t^2-\mu_{Y2}|\mathcal{F}_{t-m})| &\leq E\Bigg(|(c_1^2)^m y_{t-m}^2| \\
&\quad +|-\mu_{Y2}+\sum_{j=0}^{m-1}(c_1^2)^j\Big\{r_y^2\exp(a_1^{m-j}w_{t-m})\exp(\frac{1}{2}\sum_{l=0}^{m-j-1}a_1^{2l}r_w^2)\Big\}|\Bigg) \\
&\leq (c_1^2)^m E y_{t-m}^2 \\
&\quad +|-\mu_{Y2}+\sum_{j=0}^{m-1}(c_1^2)^j r_y^2\exp(\frac{a_1^{2(m-j)}}{2}\frac{r_w^2}{1-a_1^2})\exp(\frac{r_w^2}{2}\sum_{l=0}^{m-j-1}a_1^{2l})|\ .
\end{aligned}
\tag{A.62}
$$

Since

$$
\begin{aligned}
\lim_{m\to\infty} r_y^2\sum_{j=0}^{m-1}(c_1^2)^j\exp(\frac{a_1^{2(m-j)}}{2}\frac{r_w^2}{1-a_1^2})\exp(\frac{r_w^2}{2}\sum_{l=0}^{m-j-1}a_1^{2l}) &= \lim_{m\to\infty}\Bigg[r_y^2\sum_{j=0}^{m-1}(c_1^2)^j\exp(\frac{a_1^{2(m-j)}}{2}\frac{r_w^2}{1-a_1^2}) \\
&\quad \exp(\frac{r_w^2}{2}\frac{1-a_1^{2(m-j)}}{1-a_1^2})\Bigg] \\
=\lim_{m\to\infty} r_y^2\sum_{j=0}^{m-1}(c_1^2)^j\exp\Bigg[\frac{r_w^2}{2}\frac{a_1^{2(m-j)}}{1-a_1^2}+\frac{r_w^2}{2}\frac{(1-a_1^{2(m-j)})}{1-a_1^2}\Bigg] &= \lim_{m\to\infty}\sum_{j=0}^{m-1}(c_1^2)^j r_y^2\exp(\frac{r_w^2}{2(1-a_1^2)}) \\
&= \frac{1}{1-c_1^2}r_y^2\exp(\frac{r_w^2}{2(1-a_1^2)}) = \frac{\mu_2}{1-c_1^2}=\mu_{Y2}
\end{aligned}
\tag{A.63}
$$

we get

$$E|E(y_t^2-\mu_{Y2}|\mathcal{F}_{t-m})|\leq \eta_t\xi_m$$

with $\eta_t=1,\ \forall t$, and

$$\xi_m=(c_1^2)^m\mu_{Y2}+|-\mu_{Y2}+r_y^2\sum_{j=0}^{m-1}(c_1^2)^j\exp(\frac{a_1^{2(m-j)}}{2}\frac{r_w^2}{1-a_1^2})\exp(\frac{r_w^2}{2}\sum_{l=0}^{m-j-1}a_1^{2l})| \quad ,\forall\, m, \tag{A.64}$$

with $\lim_{m\to\infty}\xi_m = 0$ and $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\eta_t = 1 < \infty$. Thus, the process $\{y_t^2 - \mu_{Y2}\ , t\in\mathbb{N}\}$ is a L_1-mixingale w.r.t. the subfields $\mathcal{F}_t\ , t\in N$.

We shall show now that the sequence $\{y_t^2-\mu_{Y2}\ , t\in\mathbb{N}\}$ is uniformly integrable. By the $\tilde{c}_r-$inequality [see Loève (1963, p.155)], we get:

$$E|y_t^2-\mu_{Y2}|^2 \ \leq\ 2(E|y_t^2|^2+E|\mu_{Y2}|^2)$$

$$= \quad 2\mu_{Y4} + 2\mu_{Y2}^2 \; , \tag{A.65}$$

where

$$\begin{aligned} \mu_{Y4} \equiv Ey_t^4 &= E[c_1 y_{t-1} + u_t(c_1)]^4 \\ &= E[c_1^4 y_{t-1}^4 + 4c_1^3 y_{t-1}^3 u_t(c_1) + 6c_1^2 y_{t-1}^2 u_t^2(c_1) + 4c_1 y_{t-1} u_t^3(c_1) + u_t^4(c_1)] \end{aligned} \tag{A.66}$$

which yields, using the fact that the odd moments of $u_t(c_1)$ are zero and Assumption **2.3**

$$\begin{aligned} Ey_t^4 &= c_1^4 Ey_{t-1}^4 + 6c_1^2 Ey_{t-1}^2 Eu_t^2(c_1) + Eu_t^4(c_1) \\ &= \frac{1}{1-c_1^4}(\mu_4(\theta) + 6c_1^2 \mu_{Y2}\mu_2(\theta)) < \infty \; . \end{aligned} \tag{A.67}$$

Hence, $E|y_t^2 - \mu_{Y2}|^2$ above is finite and is equal to $E|y_t^2 - \mu_{Y2}|^{1+\theta} < \infty$, with $\theta = 1$ from which it follows that $\lim_{M\to\infty} E(|y_t^2 - \mu_{Y2}| 1_{|y_t^2 - \mu_{Y2}| \geq M}) = 0$ [see James Davidson (1994, p.190, Theorem 12.10)]. This holds for all $t \in \mathbb{N}$. We can then say that the process $\{y_t^2 - \mu_{Y2} \; , t \in \mathbb{N}\}$ is uniformly integrable.

Finally, we apply the L.L.N. for L_1-mixingales [see Hamilton (1994, p.191, proposition 7.6)] to the process $\{y_t^2 - \mu_{Y2} \; , t \in \mathbb{N}\}$ to obtain:

$$\frac{1}{T}\sum_{t=1}^{T}(y_t^2 - \mu_{Y2}) \xrightarrow{P} 0$$

or, equivalently,

$$\frac{1}{T}\sum_{t=1}^{T} y_t^2 \xrightarrow{P} \mu_{Y2} \; . \tag{A.68}$$

The process $\{u_t(c_1)y_{t-1} \; , t \in \mathbb{N}\}$ is clearly a martingale difference sequence (m.d.s.) w.r.t. the subfields $\mathcal{F}_t'$. Hence, by the Central Limit Theorem (C.L.T.) for m.d.s. [see Hamilton (1994, p.193, Proposition 7.8)], we get:

$$\begin{aligned} \sqrt{T}(\hat{c}_T - c_1) &= \left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2\right)^{-1} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}u_t \\ &\underset{T\to\infty}{\xrightarrow{D}} (\mu_{Y2})^{-1} Z \end{aligned} \tag{A.69}$$

with $Z \sim N(0, \frac{\mu_2^2}{1-c_1^2})$, hence,

$$\sqrt{T}(\hat{c}_T - c_1) \xrightarrow{D} N(0, 1 - c_1^2) \ . \tag{A.70}$$

$\square$

Proof of Lemma 3.4

a) To show (3.21), we examine the uniform integrability of the process. We have:

$$\begin{aligned} E|u_t y_{t-1}|^r &= E|r_y z_t \exp(w_t/2) y_{t-1}|^r \\ &= E(|z_t|^r |r_y|^r \exp(\frac{r}{2} w_t) |y_{t-1}|^r) \\ &= E|z_t|^r |r_y|^r E(\exp(\frac{r}{2} w_t)) E|y_{t-1}|^r \ . \end{aligned} \tag{A.71}$$

For a standard gaussian variable, it is known that :

$$E|z_t|^{2n} = (2n-1)!! \tag{A.72}$$

where $(2n-1)!! = 1 \times 3 \times 5 \times \ldots \times (2n-1)$ and

$$E|z_t|^{2n+1} = \sqrt{\frac{2}{\pi}} 2^n n! \tag{A.73}$$

[see Gradshteyn and Ryzhik (1980, p.337, formula 3.461.3)]. Since

$$y_{t-1} = c_1 y_{t-2} + \exp(w_{t-1}/2) r_y z_{t-1} \ ,$$

we have:

$$|y_{t-1}|^r = |c_1 y_{t-2} + \exp(w_{t-1}/2) r_y z_{t-1}|^r \ . \tag{A.74}$$

Using the $\tilde{c}_r-$inequality [see Loeve (1963, p.155)], we can say that:

$$\begin{aligned} E|y_{t-1}|^r &= E|c_1 y_{t-2} + \exp(w_{t-1}/2) r_y z_{t-1}|^r \\ &\le \tilde{c}_r \Big[E|c_1 y_{t-2}|^r + E|\exp(w_{t-1}/2) r_y z_{t-1}|^r \Big] \\ &= \tilde{c}_r \Big[|c_1|^r E|y_{t-2}|^r + |r_y|^r E|z_{t-1}|^r E \exp(\frac{r}{2} w_{t-1}) \Big] \ , \end{aligned} \tag{A.75}$$

where

$$\tilde{c}_r = 1 \quad \text{for } 0 < r \leq 1 \ ,$$
$$\tilde{c}_r = 2^{r-1} \quad \text{for } r > 1 \ .$$

By Assumption **2.3**, we have $E|y_{t-1}|^r = E|y_{t-2}|^r$, so that:

$$E|y_{t-1}|^r \quad \leq \quad \frac{\tilde{c}_r}{1 - \tilde{c}_r|c_1|^r}|r_y|^r E|z_{t-1}|^r \exp(\frac{r^2}{8}\frac{r_w^2}{1-a^2}) \equiv K_r < \infty \qquad \text{(A.76)}$$

with $1 - \tilde{c}_r|c_1|^r \neq 0$, and where $E|z_t|^r$ is given by equations (A.72) and (A.73). In particular, for $r = 1$,

$$E|y_{t-1}| \leq \frac{1}{1-|c_1|}|r_y|\sqrt{\frac{2}{\pi}}\exp(\frac{1}{8}\frac{r_w^2}{1-a_1^2}) \equiv K_1$$

Thus, we can say that:

$$E|y_{t-1}|^r \leq K_r < \infty \ ,$$

for r finite. At this step, it is now possible to compute equation (A.71) which becomes

$$E|u_t y_{t-1}|^r \leq E|z_t|^r |r_y|^r \exp(\frac{r^2}{8}\frac{r_w^2}{1-a_1^2})K_r \equiv B_r < \infty \qquad \text{(A.77)}$$

and this holds for any r finite. $E|u_t y_{t-1}|^r = E|u_t y_{t-1}|^{1+\theta}$ with $\theta = (r-1) > 0$ i.e. $r > 1$ from which it follows that:

$$\lim_{M\to\infty} E(|u_t(c_1)y_{t-1}| 1_{|u_t(c_1)y_{t-1}| \geq M}) = 0$$

[see Davidson, (1994, p.190 Theorem 12.10)]. And this holds for all $t \in \mathbb{N}\backslash\{0\}$. Thus, the collection $\{u_t(c_1)y_{t-1},\ t \in \mathbb{N}\backslash\{0\}\}$ is uniformly integrable.

Second, the process $\{u_t(c_1)y_{t-1},\ t \in \mathbb{N}\backslash\{0\}\}$ which is a m.d.s. w.r.t. $\mathcal{F}_t$ can be described as a L_1-mixingale w.r.t. $\mathcal{F}_t$ with $\xi_0 = 1$, $\xi_m = 0\ , m \geq 1$ and set $\eta_t = E|u_t(c_1)y_{t-1}|$ which corresponds to equation (A.77) with $r = 1$ from which it follows that $\eta_t \leq B_1 < \infty$ or:

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \eta_t \leq \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} B_1 = B_1 < \infty \ .$$

Finally, by the L.L.N. for L_1-mixingale [see Hamilton (1994, p.191, Proposition 7.6)] we conclude that :

$$\frac{1}{T}\sum_{t=1}^{T}(u_t(c_1)y_{t-1}) \xrightarrow{P} E(u_t(c_1)y_{t-1}) = 0 \ . \tag{A.78}$$

b) To prove (3.22), we proceed in a similar way. Clearly, the process $\{u_t(c_1)^3 y_{t-1}\} = \{v_t(\theta)^3 y_{t-1}\}$ is a m.d.s. w.r.t. $\mathcal{F}_t = \sigma(s_t, s_{t-1}, \ldots)$ where $s_t = (y_t, w_t)'$ since:

$$\begin{aligned} E[u_t(c_1)^3 y_{t-1}|\mathcal{F}_{t-1}] &= y_{t-1}E[u_t(c_1)^3|\mathcal{F}_{t-1}] \\ &= y_{t-1}E[v_t(\theta)^3|\mathcal{F}_{t-1}] \\ &= y_{t-1}E[\exp(\frac{3}{2}w_t)r_y^3 z_t^3|\mathcal{F}_{t-1}] \\ &= y_{t-1}r_y^3 E(z_t^3|\mathcal{F}_{t-1})E[\exp(\frac{3}{2}w_t)|\mathcal{F}_{t-1}] \\ &= y_{t-1}r_y^3 E(z_t^3)E[\exp(\frac{3}{2}w_t)] \\ &= 0 \ , \end{aligned} \tag{A.79}$$

from which it follows that $E[u_t(c_1)^3 y_{t-1}] = 0$, $\forall t \in \mathbb{N}\backslash\{0\}$.

A little algebra yields for a finite positive integer r that:

$$\begin{aligned} E|u_t(c_1)^3 y_{t-1}|^r &= E|z_t|^{3r}|r_y|^{3r}E[\exp(\frac{3r}{2}w_t)]E|y_{t-1}|^r \\ &\leq E|z_t|^{3r}|r_y|^{3r}\exp(\frac{9r^2}{8}\frac{r_w^2}{1-a_1^2})K_r \equiv B_r < \infty \end{aligned} \tag{A.80}$$

according to equations (A.76), (A.72) and (A.73). Further, $E|u_t(c_1)^3 y_{t-1}|^r = E|u_t(c_1)^3 y_{t-1}|^{1+\theta} < \infty$ for $\theta = r-1 > 0$ from which it follows that

$$\lim_{M\to\infty} E\Bigg(|u_t(c_1)^3 y_{t-1}|1_{|u_t(c_1)^3 y_{t-1}|\geq M}\Bigg) = 0$$

[see James Davidson, (1994), p.190, Theorem 12.10]. And this holds for all $t \in \mathbb{N}\backslash\{0\}$. Thus the collection $\{u_t(c_1)^3 y_{t-1} \ , t \in \mathbb{N}\backslash\{0\}\}$ is uniformly integrable.

Second, the process $\{u_t(c_1)^3 y_{t-1} \ , t \in \mathbb{N}\backslash\{0\}\}$ which is a m.d.s. w.r.t. $\mathcal{F}_t$ can be described as a L_1-mixingale w.r.t. $\mathcal{F}_t$ with $\xi_0 = 1$, and $\xi_m = 0$ for $m \geq 1$. Setting

$\eta_t = E|u_t(c_0)^3 y_{t-1}|$, we can see at the light of equation (A.80) with $r = 1$ that

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \eta_t \leq \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} B_1 = B_1 < \infty \ .$$

Finally, by the L.L.N. for L_1-mixingale we establish that:

$$\frac{1}{T}\sum_{t=1}^{T} u_t(c_1)^3 y_{t-1} \xrightarrow{P} E(u_t(c_1)^3 y_{t-1}) = 0 \ . \tag{A.81}$$

c) To prove (3.23), we need to check that the conditions required by the L.L.N. for L_1-mixingales are met for this process. Note that $E(u_t^2(c_1)u_{t-1}(c_1)y_{t-2}) = E(y_{t-2})E(u_t^2 u_{t-1}) = 0$ since $E(y_{t-2}) = 0$. The proof is structured as follows. In a first step we show that the process $u_t^2(c_1)u_{t-1}(c_1)y_{t-2}$ is a L_1-mixingale. To do so, we need to show that

$$E|E[u_t^2(c_1)u_{t-1}(c_1)y_{t-2}|\mathcal{F}_{t-m}]| \leq \eta_t \xi_m$$

with $\lim_{m\to\infty}\xi_m = 0$. Let us first compute $E[u_t^2(c_1)u_{t-1}(c_1)y_{t-2}|\mathcal{F}_{t-m}]$ for $m \geq 3$, i.e.

$$\begin{aligned}
E[u_t^2(c_1)u_{t-1}(c_1)y_{t-2}|\mathcal{F}_{t-m}] &= E[u_t^2 u_{t-1}(c_1^{m-2}y_{t-m} + c_1^{m-3}u_{t-m+1} + c_1^{m-4}u_{t-m+2} + \ldots \\
&\quad + c_1 u_{t-3} + u_{t-2})|\mathcal{F}_{t-m}] \\
&= E[c_1^{m-2}y_{t-m}u_t^2 u_{t-1} + c_1^{m-3}u_t^2 u_{t-1}u_{t-m+1} + c_1^{m-4}u_t^2 u_{t-1}u_{t-m+2} \\
&\quad + \ldots + c_1 u_t^2 u_{t-1}u_{t-3} + u_t^2 u_{t-1}u_{t-2}|\mathcal{F}_{t-m}] \\
&= 0
\end{aligned} \tag{A.82}$$

since $E(z_{t-1}|\mathcal{F}_{t-m}) = E(z_{t-1}) = 0$.

Similarly, we also have that $E[u_t^2(c_1)u_{t-1}(c_1)y_{t-2}|\mathcal{F}_{t-2}] = 0$. Therefore,

$$E|E[u_t^2(c_1)u_{t-1}(c_1)y_{t-2}|\mathcal{F}_{t-m}]| = 0 \ , \quad m \geq 2$$

Now for $m = 1$ we have:

$$\begin{aligned}
E[u_t^2 u_{t-1}y_{t-2}|\mathcal{F}_{t-1}] &= u_{t-1}y_{t-2}E[r_y^2 z_t^2 \exp(a_1 w_{t-1} + r_w v_t)|\mathcal{F}_{t-1}] \\
&= u_{t-1}y_{t-2}r_y^2 E(z_t^2|\mathcal{F}_{t-1})\exp(a_1 w_{t-1})E[\exp(r_w v_t)|\mathcal{F}_{t-1}] \\
&= u_{t-1}y_{t-2}r_y^2 \exp(a_1 w_{t-1})E(z_t^2)E[\exp(r_w v_t)] \\
&= u_{t-1}y_{t-2}r_y^2 \exp(a_1 w_{t-1})\exp(\frac{1}{2}r_w^2) \ .
\end{aligned} \tag{A.83}$$

Then, we have:

$$
\begin{aligned}
E|E[u_t^2 u_{t-1} y_{t-2}|\mathcal{F}_{t-1}]| &= E|r_y z_{t-1}\exp(w_{t-1}/2) y_{t-2} r_y^2 \exp(a_1 w_{t-1})\exp(\frac{1}{2}r_w^2)| \\
&= E\Bigg(\exp[(a_1+\frac{1}{2})w_{t-1}]\exp(\frac{1}{2}r_w^2)|r_y^3||z_{t-1}||y_{t-2}|\Bigg) \\
&= |r_y^3|\exp(\frac{1}{2}r_w^2)\sqrt{\frac{2}{\pi}}E|y_{t-2}|\exp[\frac{(a_1+\frac{1}{2})^2}{2}\frac{r_w^2}{1-a_1^2}] \\
&\leq |r_y^3|\exp(\frac{1}{2}r_w^2)\sqrt{\frac{2}{\pi}}K_1\exp[\frac{(a_1+\frac{1}{2})^2}{2}\frac{r_w^2}{1-a_1^2}] \equiv B
\end{aligned}
\tag{A.84}
$$

recalling that $E|y_{t-2}| \leq K_1$ and $E|z_{t-1}| = \sqrt{\frac{2}{\pi}}$. A similar calculation yields also that

$$
E|E[u_t^2 u_{t-1} y_{t-2}|\mathcal{F}_t]| = E|u_t^2 u_{t-1} y_{t-2}| \leq B\ .
$$

Then, the process $\{u_t^2 u_{t-1} y_{t-2}\}$ is a L_1-mixingale with $\eta_t = E|u_t^2 u_{t-1} y_{t-2}|$, $\forall t$, $\xi_m = 1$ for $m = 0, 1$, and $\xi_m = 0$ for $m \geq 2$, and $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\eta_t \leq B < \infty$.

On the other hand, we show that the process $\{u_t^2 u_{t-1} y_{t-2}\}$ is uniformly integrable. To do this, we shall compute $E|u_t^2 u_{t-1} y_{t-2}|^r$ for $r = 1, 2, 3, \ldots$.

$$
\begin{aligned}
E|u_t^2 u_{t-1} y_{t-2}|^r &= E|r_y^2 z_t^2 \exp(w_t) r_y z_{t-1}\exp(w_{t-1}/2) y_{t-2}|^r \\
&= E|r_y^3 z_t^2 z_{t-1} y_{t-2}\exp(a_1 w_{t-1} + r_w v_t)\exp(w_{t-1}/2)|^r \\
&= E\Bigg(|r_y^3||z_t^2||z_{t-1}||y_{t-2}|\exp[(a_1+\frac{1}{2})w_{t-1}]\exp(r_w v_t)\Bigg)^r \\
&= |r_y^3|^r E z_t^{2r} E|z_{t-1}|^r E|y_{t-2}|^r \exp[\frac{r^2(a_1+\frac{1}{2})^2}{2}\frac{r_w^2}{1-a_1^2}]\exp(\frac{r^2 r_w^2}{2}) \\
&\leq |r_y^3|^r \frac{(2r)!}{2^r r!}E|z_{t-1}|^r K_r \exp[\frac{r^2(a_1+\frac{1}{2})^2}{2}\frac{r_w^2}{1-a_1^2}]\exp(\frac{r^2 r_w^2}{2}) \\
&< \infty \quad \text{for } r < \infty\ .
\end{aligned}
\tag{A.85}
$$

Noting that:

$$
E|u_t^2 u_{t-1} y_{t-2}|^r = E|u_t^2 u_{t-1} y_{t-2}|^{1+\theta} < \infty
$$

for $\theta = (r-1) > 0$ i.e. $r > 1$, r a finite integer, then it follows that

$$
\lim_{M\to\infty} E(|u_t^2 u_{t-1} y_{t-2}| 1_{|u_t^2 u_{t-1} y_{t-2}| \geq M}) = 0\ ,
$$

[see Davidson (1994, p.190, Theorem 12.10)]. And this holds for any $t \in N\backslash\{0,1\}$. Then, the collection $\{u_t^2 u_{t-1} y_{t-2}\ , t \in \mathbb{N}\backslash\{0,1\}\}$ is uniformly integrable.

Finally, by the L.L.N. for L_1-mixingales we deduce that:

$$\frac{1}{T}\sum_{t=1}^{T} u_t^2 u_{t-1} y_{t-2} \xrightarrow{P} E u_t^2 u_{t-1} y_{t-2} = 0\ .$$

d) Finally, to prove (3.24), we need to show that the process $\{u_t u_{t-1}^2 y_{t-1}\}$ is on one hand a L_1-mixingale w.r.t. the subfield $\mathcal{F}_t$ and on the other hand uniformly integrable for the L.L.N for L_1-mixingales to hold. Let us show first that the process $\{u_t u_{t-1}^2 y_{t-1}\}$ is a L_1-mixingale w.r.t. $\mathcal{F}_t$. More precisely, it is a m.d.s. w.r.t. $\mathcal{F}_t$ since:

$$\begin{aligned}
E[u_t(c_1)u_{t-1}^2(c_1)y_{t-1}|\mathcal{F}_{t-1}] &= u_{t-1}^2(c_1)y_{t-1}E[u_t(c_1)|\mathcal{F}_{t-1}] \\
&= u_{t-1}^2(c_1)y_{t-1}E[r_y z_t \exp(w_t/2)|\mathcal{F}_{t-1}] \\
&= r_y u_{t-1}^2(c_1)y_{t-1}E[z_t|\mathcal{F}_{t-1}]E[\exp(w_t/2)\mathcal{F}_{t-1}] \\
&= 0\ , \qquad (A.86)
\end{aligned}$$

since $z_t \overset{iid}{\sim} N(0,1)$. Hence, we deduce that $E(u_t(c_1)u_{t-1}^2(c_1)y_{t-1}) = 0$. Therefore, the process $\{u_t u_{t-1}^2 y_{t-1}\}$ which is a m.d.s. w.r.t. $\mathcal{F}_t$ can be described as a specific L_1-mixingale w.r.t. $\mathcal{F}_t$ with $\xi_0 = 1$ and $\xi_m = 0$ for $m \geq 1$.

On the second hand we show that this process is uniformly integrable. Using once again the $\tilde{c}_r$-inequality [see Loève (1963, p.155)] we can state that:

$$\begin{aligned}
E|u_t(c_1)u_{t-1}^2(c_1)y_{t-1}|^r &= E|r_y^3 z_t z_{t-1}^2 \exp(\frac{w_t}{2} + w_{t-1})c_1 y_{t-2} + r_y^4 z_t z_{t-1}^3 \exp(\frac{w_t + 3w_{t-1}}{2})|^r \\
&\leq \tilde{c}_r \Bigg\{ |c_1|^r E|y_{t-2}|^r |r_y^3|^r E|z_t|^r E|z_{t-1}^2|^r \exp\left[\frac{r^2}{8}(2+a_1)^2 \frac{r_w^2}{1-a_1^2}\right] \\
&\quad \exp(\frac{r^2 r_w^2}{8}) + |r_y^4|^r E|z_t|^r E|z_{t-1}^3|^r \exp\left[\frac{r^2}{8}(3+a_1)^2 \frac{r_w^2}{1-a_1^2}\right] \\
&\quad \exp(\frac{r^2 r_w^2}{8}) \Bigg\} \\
&\leq \tilde{c}_r \Bigg\{ |c_1|^r K_r |r_y|^{3r} \gamma_r \frac{(2r)!}{2^r r!} \exp\left[\frac{r^2}{8}(2+a_1)^2 \frac{r_w^2}{1-a_1^2}\right] \exp(\frac{r^2 r_w^2}{8}) \\
&\quad + |r_y|^{4r} \gamma_r \gamma_{3r} \exp\left[\frac{r^2}{8}(3+a_1)^2 \frac{r_w^2}{1-a_1^2}\right] \exp(\frac{r^2 r_w^2}{8}) \Bigg\} \\
&\equiv B_r < \infty \qquad (A.87)
\end{aligned}$$

where it has been shown earlier that $E|y_{t-2}|^r \leq K_r$ by equation (A.76), that $\gamma_r \equiv$

$E|z_t|^r$, $\gamma_{3r} \equiv E|z_{t-1}|^{3r}$, and $E|z_{t-1}^2|^r = Ez_{t-1}^{2r} = \frac{(2r)!}{2^r r!}$ since $z_t \sim N(0,1)$, and γ_r and γ_{3r} are given by equations (A.72), (A.73). Noting that $E|u_t(c_1)u_{t-1}^2(c_1)y_{t-1}|^r = E|u_t(c_1)u_{t-1}^2(c_1)y_{t-1}|^{1+\theta} < \infty$ with $\theta = r-1 > 0$, i.e. $1 < r < \infty$, it follows that

$$\lim_{M\to\infty} E\left(|u_t(c_1)u_{t-1}^2(c_1)y_{t-1}| 1_{|u_t(c_1)u_{t-1}^2(c_1)y_{t-1}|\geq M} \right) = 0 \; .$$

And this holds for all $t \in \mathbb{N}\backslash\{0\}$. Hence, the collection $\{u_t(c_1)u_{t-1}^2(c_1)y_{t-1} \; , \forall \; t \in \mathbb{N}\backslash\{0\}\}$ is uniformly integrable. Besides, taking $\eta_t = E|u_t(c_1)u_{t-1}^2(c_1)y_{t-1}|$ which corresponds to equation (A.87) with $r = 1$ yields

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \eta_t \leq \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} B_1 = B_1 < \infty \; .$$

We can finally apply the L.L.N. for L_1-mixingales to state that:

$$\frac{1}{T}\sum_{t=1}^{T} u_t(c_1)u_{t-1}^2(c_1)y_{t-1} \xrightarrow{P} E(u_t(c_1)u_{t-1}^2(c_1)y_{t-1}) = 0 \; . \tag{A.88}$$

$\square$

Proof of Proposition 3.5

The asymptotic equivalence of the function:

$$\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta)) = \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^2 - \mu_2(\theta)) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^4 - \mu_4(\theta)) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^2\hat{u}_{t-1}^2 - \mu_{2,2}(1)(\theta)) \end{pmatrix} \tag{A.89}$$

to $\sqrt{T}(\bar{g}_T(\theta) - \mu(\theta))$ will be shown component by component.

1. **The component** $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^2 - \mu_2(\theta))$

   Recall that $u_t^2(c) \equiv (y_t - cy_{t-1})^2$ and $\hat{u}_t^2 \equiv u_t^2(\hat{c}_T)$. Noting that:

$$\begin{aligned} u_t^2(\hat{c}_T) - u_t^2(c_1) &= (\hat{c}_T^2 - c_1^2)y_{t-1}^2 - 2(\hat{c}_T - c_1)y_t y_{t-1} \\ &= -2(\hat{c}_T - c_1)y_{t-1}u_t(c_1) + y_{t-1}^2(\hat{c}_T - c_1)^2 \; , \end{aligned} \tag{A.90}$$

   we deduce after aggregation:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^2(\hat{c}_T)-\mu_2(\theta)) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^2(c_1)-\mu_2(\theta))-\sqrt{T}(\hat{c}_T-c_1)\frac{2}{T}\sum_{t=1}^{T} y_{t-1}u_t(c_1)$$
$$+\sqrt{T}(\hat{c}_T-c_1)(\hat{c}_T-c_1)\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \,. \quad \text{(A.91)}$$

By Lemmas **3.3**,**3.4** the last two terms of the right-hand side of equation (A.91) are both $o_p(1)$ variables. Hence, equation (A.91) is equivalent to

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^2-\mu_2(\theta))\#\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t(c_1)^2-\mu_2(\theta))$$

asymptotically.

2. **The component** $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^4-\mu_4(\theta))$ Noting that:

$$\begin{aligned} u_t^4(\hat{c}_T)-u_t^4(c_1) &= -4y_{t-1}u_t^3(c_1)(\hat{c}_T-c_1)+6y_{t-1}^2u_t^2(c_1)(\hat{c}_T-c_1)^2-4y_{t-1}^3u_t(c_1)(\hat{c}_T-c_1)^3 \\ &\quad +y_{t-1}^4(\hat{c}_T-c_1)^4 \end{aligned} \quad \text{(A.92)}$$

yields after aggregation:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^4(\hat{c}_T)-\mu_4(\theta)) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t(c_1)^4-\mu_4(\theta))+R_T \quad \text{(A.93)}$$

where

$$\begin{aligned} R_T &\equiv -\frac{4}{T}\sum_{t=1}^{T} y_{t-1}u_t(c_1)^3\sqrt{T}(\hat{c}_T-c_1)+6\sqrt{T}(\hat{c}_T-c_1)^2\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2u_t^2(c_1) \\ &\quad -4\sqrt{T}(\hat{c}_T-c_1)^3\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^3u_t(c_1)+\sqrt{T}(\hat{c}_T-c_1)^4\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^4 \,. \end{aligned} \quad \text{(A.94)}$$

We shall show here that R_T is an $o_p(1)$-variable. To do so, let us first focus on the second component of R_T. Set

$$X_T = \frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2u_t^2(c_1) \,,$$

so $|X_T| = X_T$ and

$$E(X_T) = \frac{1}{T}\sum_{t=1}^{T} E[y_{t-1}^2 u_t^2(c_1)] = \frac{1}{T}\sum_{t=1}^{T} E(y_{t-1}^2)E(u_t^2(c_1)) = \mu_{Y2}\mu_2(\theta) < \infty \; . \tag{A.95}$$

Hence by Markov inequality we have:

$$P[X_T \geq \epsilon] \leq \frac{E(X_T)}{\epsilon} < \infty \; \forall \epsilon > 0.$$

Now for the third component of R_T, set:

$$X_T = -\frac{4}{T}\sum_{t=1}^{T} y_{t-1}^3 u_t(c_1) \; ,$$

so that

$$\begin{aligned} |X_T| &\leq \frac{4}{T}\sum_{t=1}^{T} |y_{t-1}^3 u_t(c_1)| \\ &= \frac{4}{T}\sum_{t=1}^{T} |y_{t-1}^3 \exp(w_t/2)|r_y||z_t| \end{aligned} \tag{A.96}$$

which yields:

$$E|X_T| \leq \frac{4}{T}\sum_{t=1}^{T} E|y_{t-1}|^3 |r_y| E|z_t| E\exp(w_t/2) < \infty \; , \tag{A.97}$$

by equations (A.73), (A.76) and by $E\exp(w_t/2) = \exp[r_w^2/(8(1-a_1^2))] < \infty$. Once again by Markov inequality, we have:

$$P[|X_T| \geq \epsilon] \leq \frac{E|X_T|}{\epsilon} < \infty \; \forall \epsilon > 0.$$

Similarly, for the fourth component of R_T, set:

$$X_T = \frac{1}{T}\sum_{t=1}^{T} y_{t-1}^4 \; \geq 0$$

so that $|X_T| = X_T$ and $EX_T = \frac{1}{T}\sum_{t=1}^{T} y_{t-1}^4 = \mu_{Y4} < \infty$, we can say by Markov inequality that:

$$P[X_T \geq \epsilon] \leq \frac{E(X_T)}{\epsilon} < \infty \ \forall \epsilon > 0.$$

Finally, by Lemmas **3.3** and **3.4**, we can conclude that R_T is an $o_p(1)$-variable which yields that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\hat{u}_t^4 - \mu_4(\theta)) \# \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (u_t(c_1)^4 - \mu_4(\theta))$$

asymptotically.

3. **The component** $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\hat{u}_t^2 \hat{u}_{t-1}^2 - \mu_{2,2}(1|\theta))$

   Noting that:

$$\begin{aligned} u_t^2(\hat{c}_T) u_{t-1}^2(\hat{c}_T) - u_t^2(c_1) u_{t-1}^2(c_1) &= -2(\hat{c}_T - c_1)[u_t(c_1) u_{t-1}^2(c_1) y_{t-1} + u_t^2(c_1) u_{t-1}(c_1) y_{t-2}] \\ &+ (\hat{c}_T - c_1)^2 [y_{t-1}^2 u_{t-1}^2(c_1) + y_{t-2}^2 u_t^2(c_1) + 4 y_{t-1} y_{t-2} u_t(c_1) u_{t-1}(c_1)] \\ &- 2(\hat{c}_T - c_1)^3 [y_{t-1}^2 y_{t-2} u_{t-1}(c_1) + y_{t-2}^2 y_{t-1} u_t(c_1)] + (\hat{c}_T - c_1)^4 y_{t-1}^2 y_{t-2}^2 \end{aligned} \tag{A.98}$$

   yields after aggregation

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (u_t^2(\hat{c}_T) u_{t-1}^2(\hat{c}_T) - \mu_{2,2}(1|\theta)) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (u_t(c_1)^2 u_{t-1}(c_1)^2 - \mu_{2,2}(1|\theta)) + R_T \tag{A.99}$$

   where

$$\begin{aligned} R_T &= -2\sqrt{T}(\hat{c}_T - c_1) \frac{1}{T} \sum_{t=1}^{T} [u_t(c_1) u_{t-1}^2(c_1) y_{t-1} + u_t^2(c_1) u_{t-1}(c_1) y_{t-2}] \\ &+ \sqrt{T}(\hat{c}_T - c_1)^2 \frac{1}{T} \sum_{t=1}^{T} [y_{t-1}^2 u_{t-1}^2(c_1) + y_{t-2}^2 u_t^2(c_1) + 4 y_{t-1} y_{t-2} u_t(c_1) u_{t-1}(c_1)] \\ &-2\sqrt{T}(\hat{c}_T - c_1)^3 \frac{1}{T} \sum_{t=1}^{T} [y_{t-1}^2 y_{t-2} u_{t-1}(c_1) + y_{t-2}^2 y_{t-1} u_t(c_1)] + \sqrt{T}(\hat{c}_T - c_1)^4 \frac{1}{T} \sum_{t=1}^{T} y_{t-1}^2 y_{t-2}^2 \end{aligned} \tag{A.100}$$

Let us focus on the following components of R_T :

$$\frac{1}{T}\sum_{t=1}^{T}\Big\{y_{t-1}^2u_{t-1}^2(c_1)+y_{t-2}^2u_t^2(c_1)+4y_{t-1}y_{t-2}u_t(c_1)u_{t-1}(c_1)+y_{t-1}^2y_{t-2}u_{t-1}(c_1)$$
$$+y_{t-2}^2y_{t-1}u_t(c_1)+y_{t-1}^2y_{t-2}^2\Big\} \quad \text{(A.101)}$$

If we can show that the expectation of the absolute value of each one of these components is finite, then by Markov inequality we will be able to state that these quantities are bounded in probability and then conclude that R_T is an $o_p(1)$-variable by Lemmas **3.3** and **3.4**. In this aim, let us compute the expectation of the absolute value of corresponding quantities. Set:

$$X_T=\frac{1}{T}\sum_{t=1}^{T}y_{t-1}^2u_{t-1}^2(c_1)\geq 0 \quad \text{(A.102)}$$

so that $|X_T|=X_T$ and

$$EX_T=\frac{1}{T}\sum_{t=1}^{T}E[(c_1y_{t-2}+u_{t-1}(c_1))^2u_{t-1}^2(c_1)]=c_1^2\mu_{Y2}\mu_2(\theta)+\mu_4(\theta)<\infty$$

by Assumptions **2.2** and **2.3**. Similarly,

$$E\left[\frac{1}{T}\sum_{t=1}^{T}y_{t-2}^2u_t^2(c_1)\right]=\mu_{Y2}\mu_2(\theta)<\infty\ .$$

Now, set:

$$X_T=\frac{4}{T}\sum_{t=1}^{T}y_{t-1}y_{t-2}u_t(c_1)u_{t-1}(c_1)$$

and replacing y_{t-1} by $c_1y_{t-2}+u_{t-1}(c_1)$ yields after taking the expectation of the absolute value :

$$\begin{aligned}E|X_T| &\leq \frac{4}{T}\sum_{t=1}^{T}\Big\{E|c_1y_{t-2}^2u_t(c_1)u_{t-1}(c_1)|+E|y_{t-2}u_{t-1}^2(c_1)u_t(c_1)|\Big\}\\ &= \frac{4}{T}\sum_{t=1}^{T}\Big\{E|c_1y_{t-2}^2|E|\exp(w_t/2)r_y^2z_tz_{t-1}\exp(w_{t-1}/2)|\end{aligned}$$

$$+E|y_{t-2}|E|\exp(w_{t-1})r_y^2 z_{t-1}^2 \exp(w_t/2) r_y z_t|\Big\} \tag{A.103}$$

where the equality comes from replacing the perturbations by their expression. Then, exploiting the log-normality of the perturbations $\exp(w_t)$, the independence property of the innovations and Assumption **2.2**, we get:

$$\begin{aligned} E|X_T| \leq \frac{4}{T}\sum_{t=1}^{T}\Big\{ & |c_1|\mu_{Y2} r_y^2 \exp[(\frac{(a_1+1)^2}{8})\frac{r_w^2}{1-a_1^2}]\exp(\frac{r_w^2}{8})E|z_t|E|z_{t-1}| \\ & + K_1|r_y|^3 \exp[1/2(1+a_1/2)^2\frac{r_w^2}{1-a_1^2}]\exp(\frac{r_w^2}{8})E|z_t|\Big\} \end{aligned} \tag{A.104}$$

where the summation symbol disappears (the quantities inside do not depend on t any more by Assumption **2.3**) and where $\mu_{Y2} = Ey_{t-2}^2$, $K_1 = E|y_{t-2}| < \infty$ by equation (A.76) and $E|z_t| < \infty$ by equation (A.73). Hence $E|X_T| < \infty$.

Now, consider:

$$X_T = \frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 y_{t-2} u_{t-1}(c_1)\ ,$$

then

$$\begin{aligned} E|X_T| &\leq \frac{1}{T}\sum_{t=1}^{T} E|y_{t-1}^2 y_{t-2} u_{t-1}(c_1)| \\ &= \frac{1}{T}\sum_{t=1}^{T} E|c_1^2 y_{t-2}^3 u_{t-1}(c_1) + 2c_1 y_{t-2}^2 u_{t-1}^2(c_1) + y_{t-2} u_{t-1}^3(c_1)| \\ &\leq \frac{1}{T}\sum_{t=1}^{T}\Big\{c_1^2 E|y_{t-2}|^3 E|u_{t-1}(c_1)| + 2|c_1|Ey_{t-2}^2 Eu_{t-1}^2(c_1) + E|y_{t-2}|E|u_{t-1}(c_1)|^3\Big\} \end{aligned} \tag{A.105}$$

where the summation symbol disappears (the quantities inside do not depend on t any more by Assumption **2.3**) and we know that

$$E|u_{t-1}(c_1)| = E|\exp(w_{t-1}/2) r_y z_{t-1}| = |r_y|\exp(\frac{r_w^2}{8(1-a_1^2)})E|z_{t-1}| < \infty$$

and

$$E|u_{t-1}(c_1)|^3 = |r_y|^3 \exp(\frac{9r_w^2}{8(1-a_1^2)})E|z_{t-1}|^3 < \infty$$

since $E|z_{t-1}|^{2n+1} < \infty$ by equation (A.73) and $E|y_{t-2}|^r = k_r < \infty$ by equation (A.76), we can deduce that $E|X_T| < \infty$.

The proof is similar for

$$X_T = \frac{1}{T}\sum_{t=1}^{T} y_{t-2}^2 y_{t-1} u_t(c_1)$$

yielding that:

$$\begin{aligned} E|X_T| \leq \frac{1}{T}\sum_{t=1}^{T}\Bigg\{ & |c_1|E|y_{t-2}|^3 \exp(\frac{r_w^2}{8(1-a_1^2)})|r_y|E|z_t| \\ & + \mu_{Y2} r_y^2 \exp[\frac{(a_1+1)^2}{8}\frac{r_w^2}{1-a_1^2}] \exp(\frac{r_w^2}{8})E|z_{t-1}|E|z_t|\Bigg\} \end{aligned} \tag{A.106}$$

which is finite by the same arguments as above. Now, if we set

$$X_T = \frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 y_{t-2}^2 \geq 0$$

then $|X_T| = X_T$ and

$$\begin{aligned} EX_T &= \frac{1}{T}\sum_{t=1}^{T} E[y_{t-1}^2 y_{t-2}^2] \\ &= \frac{1}{T}\sum_{t=1}^{T} E[(c_1 y_{t-2} + u_{t-1}(c_1))^2 y_{t-2}^2] \\ &= c_1^2 \mu_{Y4} + \mu_{Y2}\mu_2(\theta) < \infty\ . \end{aligned} \tag{A.107}$$

Therefore all the quantities appearing in equation (A.101) have a finite expectation of the absolute value, hence by Markov inequality we can say that they are bounded in probability to finally conclude by Lemmas **3.3** and **3.4** that R_T is an $o_p(1)$-variable.

Therefore, we have the asymptotic equivalence below.

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^2\hat{u}_{t-1}^2-\mu_{2,2}(1)(\theta))\#\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t(c_1)^2u_{t-1}(c_1)^2-\mu_{2,2}(1)(\theta)) \ .$$

Thus,

$$\begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^2-\mu_2(\theta)) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^4-\mu_4(\theta)) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\hat{u}_t^2\hat{u}_{t-1}^2-\mu_{2,2}(1)(\theta)) \end{pmatrix} \overset{asy}{\#} \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^2(c_1)-\mu_2(\theta)) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^4(c_1)-\mu_4(\theta)) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^2(c_1)u_{t-1}^2(c_1)-\mu_{2,2}(1)(\theta)) \end{pmatrix} \tag{A.108}$$

and from equation (3.8) we know that $u_t(c_1)=v_t(\theta)\ \forall\ t$ then we have the asymptotic equivalence

$$\sqrt{T}(\bar{g}_T(\hat{U})-\mu(\theta))\overset{asy}{\#}\sqrt{T}(\bar{g}_T(\theta)-\mu(\theta)) \ ,$$

with $\bar{g}_T(\theta)$ defined as in equation (3.18) □

Proof of Lemma 3.6

Here we derive the covariances of the components of $X_t=(X_{1t},X_{2t},X_{3t})'$ that is

$$\begin{aligned}
\gamma_1(\tau) &= Cov(X_{1t},X_{1,t+\tau})=E[(v_t^2(\theta)-\mu_2(\theta))(v_{t+\tau}^2(\theta)-\mu_2(\theta))] \\
&= E(v_t^2(\theta)v_{t+\tau}^2(\theta))-\mu_2^2(\theta) \\
&= r_y^4E\exp(w_t+w_{t+\tau})-\mu_2^2(\theta) \\
&= r_y^4\exp(\frac{r_w^2}{1-a_1^2}(1+a_1^\tau))-\mu_2^2(\theta) \\
&= \mu_2^2(\theta)[\exp(\gamma a_1^\tau)-1] \ ,
\end{aligned} \tag{A.109}$$

where $\gamma=\frac{r_w^2}{1-a_1^2}$. Similarly,

$$\begin{aligned}
\gamma_2(\tau) &= Cov(X_{2t},X_{2,t+\tau})=E[(v_t^4(\theta)-\mu_4(\theta))(v_{t+\tau}^4(\theta)-\mu_4(\theta))] \\
&= E(v_t^4(\theta)v_{t+\tau}^4(\theta))-\mu_4^2(\theta) \\
&= 9r_y^8E\exp[2(w_t+w_{t+\tau})]-\mu_4^2(\theta) \\
&= 9r_y^8\exp(4\frac{r_w^2}{1-a_1^2}(1+a_1^\tau)-\mu_4^2(\theta)) \\
&= \mu_4^2(\theta)[\exp(4\gamma a_1^\tau)-1] \ .
\end{aligned} \tag{A.110}$$

Finally,

$$\begin{aligned}
\gamma_3(\tau) &= Cov(X_{3t}, X_{3,t+\tau}) = E[(v_t^2(\theta)v_{t-1}^2(\theta) - \mu_{2,2}(1|\theta))(v_{t+\tau}^2(\theta)v_{t+\tau-1}^2(\theta) - \mu_{2,2}(1|\theta))] \\
&= E(v_t^2(\theta)v_{t-1}^2(\theta)v_{t+\tau}^2(\theta))v_{t+\tau-1}^2(\theta) - \mu_2^2(\theta) \\
&= r_y^8 E\exp(w_{t+\tau} + w_{t+\tau-1} + w_t + w_{t-1}) - \mu_2^2(\theta) \\
&= r_y^8 \exp[2(1+a_1)\gamma]\exp[\gamma(a_1^{\tau-1} + 2a_1^{\tau} + a_1^{\tau+1})] - \mu_2^2(\theta) \\
&= \mu_{2,2}^2(\theta)\{\exp[\gamma(a_1^{\tau-1} + 2a_1^{\tau} + a_1^{\tau+1})] - 1\} \\
&= \mu_{2,2}^2(\theta)\{\exp[\gamma(1+a_1)^2 a_1^{\tau-1}] - 1\} \qquad \text{(A.111)}
\end{aligned}$$

for all $\tau \geq 2$. $\square$

Proof of Proposition 3.7

In order to establish the asymptotic normality of $\sqrt{T}(\bar{g}_T(\theta) - \mu(\theta))$ we shall use a Central Limit Theorem (C.L.T) for dependent processes [see Davidson (1994, p.385, Theorem 24.5)]. For that purpose, we shall first verify the conditions under which this C.L.T holds. If we define:

$$\begin{aligned}
X_t &\equiv \begin{pmatrix} v_t^2(\theta) - \mu_2(\theta) \\ v_t^4(\theta) - \mu_4(\theta) \\ v_t^2(\theta)v_{t-1}^2(\theta) - \mu_{2,2}(1|\theta) \end{pmatrix} \\
&= g_t(\theta) - \mu(\theta)\ , \qquad \text{(A.112)}
\end{aligned}$$

$$S_T = \sum_{t=1}^{T} X_t = \sum_{t=1}^{T} g_t(\theta) - \mu(\theta)\ , \qquad \text{(A.113)}$$

and the subfields $\mathcal{F}_t = \sigma(s_t, s_{t-1}, \ldots)$ where $s_t = (y_t, w_t)'$, we need to verify three conditions, i.e.:

a) $\{X_t, \mathcal{F}_t\}$ is stationary and ergodic

b) $\{X_t, \mathcal{F}_t\}$ is a L_1-mixingale of size -1

c)
$$\limsup_{T\to\infty} T^{-1/2}E|S_T| < \infty \qquad \text{(A.114)}$$

in order to get that $T^{-1/2}S_T = \sqrt{T}(\bar{g}_T(\theta) - \mu(\theta)) \xrightarrow{D} N(0, \Omega^*)$ [see Davidson (1994, p.385, Theorem 24.5)].

a) By Propositions 5 and 17 from Carrasco,Chen (1999) we can say that

i) if $\{w_t\}$ is geometrically ergodic, then $\{(w_t, \ln|v_t|)\}$ is Markov geometrically ergodic with the same decay rate as that of $\{w_t\}$;

ii) if $\{w_t\}$ is stationary β-mixing with a certain decay rate, then $\{\ln|v_t|\}$ is β-mixing with a decay rate at least as fast as that of $\{w_t\}$.

If the initial value v_0 follows the stationary distribution, then $\{\ln|v_t|\}$ is strictly stationary β-mixing with an exponential decay rate. Since this property is preserved by any continuous transformation, $\{v_t\}$ and hence $\{v_t^k\}$ and $\{v_t^k v_{t-1}^k\}$ are strictly stationary and exponential β-mixing. We can then deduce that X_t is strictly stationary and exponential β-mixing.

b) Moreover, a mixing zero-mean process is an adapted L_1-mixingale with respect to the subfields $\mathcal{F}_t$ provided it is bounded in the L_1-norm [see Davidson (1994,p.211,Theorem 14.2)]. To see that $\{X_t\}$ is bounded in the L_1-norm, we note that:

$$\begin{aligned} E|v_t^2 - \mu_2(\theta)| &\leq E(|v_t^2| + |\mu_2(\theta)|) \\ &= 2\mu_2(\theta) < \infty \ , \end{aligned}$$

$$E|v_t^4 - \mu_4(\theta)| \leq 2\mu_4(\theta) < \infty$$

and

$$E|v_t^2 v_{t-1}^2 - \mu_{2,2}(1|\theta)| \leq 2\mu_{2,2}(1|\theta) < \infty.$$

We now need to show that the L_1-mixingale $\{X_t, \mathcal{F}_t\}$ is of size -1. Since X_t is $\beta-$mixing, it has mixing coefficients of the type $\beta_n = c\rho^n$ $, c > 0,$ $, 0 < \rho < 1$. In order to show that $\{X_t\}$ is of size -1, we need to show that its mixing coefficients $\beta_n = O(n^{-\phi})$, with $\phi > 1$.

Indeed,

$$\begin{aligned} \frac{\rho^n}{n^{-\phi}} &= n^{\phi} \exp(n \log \rho) \\ &= \exp(\phi \log n) \exp(n \log \rho) \\ &= \exp(\phi \log n + n \log \rho) \ . \end{aligned}$$

It is known [see Rudin (1976, p.57, Theorem 3.20d)] that $\lim_{n\to\infty} \phi \log n + n \log \rho = -\infty$ which yields

$$\lim_{n\to\infty} \exp(\phi \log n + n \log \rho) = 0 \ .$$

And this holds in particular with $\phi > 1$.

c) Now, the last condition to verify before applying the Central Limit Theorem for dependent processes is to show that

$$\limsup_{T\to\infty} T^{-1/2} E|S_T| < \infty$$

where

$$S_T \stackrel{def}{=} \sum_{t=1}^{T} X_t = \sum_{t=1}^{T} \begin{pmatrix} v_t^2(\theta) - \mu_2(\theta) \\ v_t^4(\theta) - \mu_4(\theta) \\ v_t^2(\theta) v_{t-1}^2(\theta) - \mu_{2,2}(1|\theta) \end{pmatrix} .$$

It is known by Cauchy-Schwarz inequality that:

$$E|T^{-1/2} S_T| \leq [E(T^{-1} S_T^2)]^{1/2} \tag{A.115}$$

so that to show equation (A.114) is equivalent to show that $\limsup_{T\to\infty} T^{-1} E(S_T^2) < \infty$.

We shall prove that:

$$\begin{aligned} \limsup_{T\to\infty} T^{-1} E(S_T^2) &= \limsup_{T\to\infty} E\left[(\frac{1}{\sqrt{T}} S_T)^2 \right] \\ &= \limsup_{T\to\infty} Var\left[\frac{1}{\sqrt{T}} S_T \right] < \infty \end{aligned} \tag{A.116}$$

i) The first component of S_T.
Set $S_{T1} = \sum_{t=1}^{T} X_{1,t}$ where $X_{1,t} \equiv v_t^2(\theta) - \mu_2(\theta)$. We compute:

$$\begin{aligned} Var\left[\frac{1}{\sqrt{T}} S_{T1} \right] &= \frac{1}{T} \left[\sum_{t=1}^{T} Var(X_{1,t}) + \sum_{\substack{t=1 \\ s\neq t}}^{T} Cov(X_{1,s}, X_{1,t}) \right] \\ &= \frac{1}{T} \left[T\gamma_1(0) + 2\sum_{\tau=1}^{T} (T-\tau)\gamma_1(\tau) \right] \\ &= \gamma_1(0) + 2\sum_{\tau=1}^{T} (1 - \frac{\tau}{T})\gamma_1(\tau) \ , \end{aligned} \tag{A.117}$$

where $\gamma \equiv r_w^2/(1-a_1^2)$. We must prove that $\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\gamma_1(\tau)$ converge as $T \to \infty$. By lemma 3.1.5 in Fuller (1976, p.112), it is sufficiant to show that $\sum_{\tau=1}^{\infty} \gamma_1(\tau)$ converge. Using the results of Lemma **3.6** we have:

$$\gamma_1(\tau) \;\; = \;\; \mu_2^2(\theta)[\exp(\gamma a_1^\tau) - 1]$$

$$\begin{aligned}
&= \mu_2^2(\theta)\left[1+\sum_{k=1}^{\infty}\frac{(\gamma a_1^\tau)^k}{k!}-1\right] \\
&= \mu_2^2(\theta)\left[\gamma a_1^\tau\sum_{k=1}^{\infty}\frac{(\gamma a_1^\tau)^{k-1}}{k!}\right] \\
&= \mu_2^2(\theta)\left[\gamma a_1^\tau\sum_{k=0}^{\infty}\frac{(\gamma a_1^\tau)^{k}}{(k+1)!}\right] \\
&\le \mu_2^2(\theta)\gamma a_1^\tau\sum_{k=0}^{\infty}\frac{(\gamma a_1^\tau)^{k}}{k!} \\
&= \mu_2^2(\theta)\gamma a_1^\tau \exp(\gamma a_1^\tau). \qquad (A.118)
\end{aligned}$$

Therefore, the series

$$\begin{aligned}
\sum_{\tau=1}^{\infty}\gamma_1(\tau) &\le \mu_2^2(\theta)\gamma\sum_{\tau=1}^{\infty}a^\tau\exp(\gamma a^\tau) \le \mu_2^2(\theta)\gamma\exp(\gamma a)\sum_{\tau=1}^{\infty}a^\tau \\
&= \mu_2^2(\theta)\frac{a\gamma\exp(\gamma a)}{1-a} < \infty \qquad (A.119)
\end{aligned}$$

converges. We deduce by Cauchy-Schwarz inequality that

$$\limsup_{T\to\infty} T^{-1/2}E|\sum_{t=1}^{T}\Big(v_t^2(\theta)-\mu_2(\theta)\Big)| < \infty \,.$$

The proof is very similar for the second component of S_T. We will skip to the third component of S_T.

ii) The third component of S_T.

Likewise, we just have to show that $\sum_{\tau=1}^{\infty}\gamma_3(\tau) < \infty$ in order to prove that $\limsup_{T\to\infty} T^{-1/2}E|\sum_{t=1}^{T}\Big(v_t^2(\theta)v_{t-1}^2(\theta)-\mu_{2,2}(\theta)\Big)| < \infty$.

By lemma **3.6** we have for all $\tau \ge 2$:

$$\begin{aligned}
\gamma_3(\tau) &= \mu_{2,2}^2(1|\theta)[\exp(\gamma(1+a_1)^2a_1^{\tau-1})-1] \\
&= \mu_{2,2}^2(1|\theta)\left\{1+\sum_{k=1}^{\infty}\frac{[\gamma(1+a_1)^2a_1^{\tau-1}]^k}{k!}-1\right\} \\
&= \mu_{2,2}^2(1|\theta)[\gamma(1+a_1)^2a_1^{\tau-1}]\sum_{k=1}^{\infty}\frac{[\gamma(1+a_1)^2a_1^{\tau-1}]^{k-1}}{k!}
\end{aligned}$$

$$
\begin{aligned}
&= \mu_{2,2}^2(1|\theta)[\gamma(1+a_1)^2 a_1^{\tau-1}] \sum_{k=0}^{\infty} \frac{[\gamma(1+a_1)^2 a_1^{\tau-1}]^k}{(k+1)!} \\
&\leq \mu_{2,2}^2(1|\theta)[\gamma(1+a_1)^2 a_1^{\tau-1}] \sum_{k=0}^{\infty} \frac{[\gamma(1+a_1)^2 a_1^{\tau-1}]^k}{k!} \\
&= \mu_{2,2}^2(1|\theta)[\gamma(1+a_1)^2 a_1^{\tau-1}] \exp[\gamma(1+a_1)^2 a_1^{\tau-1}] \ , \qquad \text{(A.120)}
\end{aligned}
$$

such that :

$$
\begin{aligned}
\sum_{\tau=1}^{\infty} \gamma_3(\tau) &\leq \gamma_3(1) + \mu_{2,2}^2(1|\theta)\gamma(1+a_1)^2 \sum_{\tau=2}^{\infty} a_1^{\tau-1} \exp[\gamma(1+a_1)^2 a_1^{\tau-1}] \\
&\leq \gamma_3(1) + \mu_{2,2}^2(1|\theta)\gamma(1+a_1)^2 \exp[\gamma(1+a_1)^2 a_1] \sum_{\tau=2}^{\infty} a_1^{\tau-1} \\
&= \gamma_3(1) + \mu_{2,2}^2(1|\theta)\gamma(1+a_1)^2 \exp[\gamma(1+a_1)^2 a_1] \sum_{\tau=1}^{\infty} a_1^{\tau} \\
&= \gamma_3(1) + \mu_{2,2}^2(1|\theta)\gamma(1+a_1)^2 \exp[\gamma(1+a_1)^2 a_1] \frac{a_1}{1-a_1} < \infty \ . \qquad \text{(A.121)}
\end{aligned}
$$

Since $\limsup_{T\to\infty} T^{-1/2} E|\sum_{t=1}^{T} X_t| < \infty$ we can therefore apply Theorem 24.5, p.385 [see Davidson (1994)] to the process $\{X_t, \mathcal{F}_t\}$ defined in equation (A.112) with $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots)$, and establish that

$$
T^{-1/2} S_T = T^{-1/2} \sum_{t=1}^{T} X_t = \sqrt{T}(\bar{g}_T(\theta) - \mu(\theta)) \xrightarrow{D} N(0, \Omega^*) \ , \qquad \text{(A.122)}
$$

where

$$
\begin{aligned}
\Omega^* &= \lim_{T\to\infty} E[(T^{-1/2} S_T)^2] \\
&= \lim_{T\to\infty} E[(\sqrt{T}(\bar{g}_T(\theta) - \mu(\theta)))^2] \\
&= \lim_{T\to\infty} E\left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (g_t(\theta) - \mu(\theta))\right]^2 \ . \qquad \text{(A.123)}
\end{aligned}
$$

□

Proof of Proposition 3.8
The method-of-moments estimator $\hat{\theta}_T(\Omega)$ is solution of the following optimization

problem:

$$\min_\theta M_T(\theta) = \min_\theta(\mu(\theta) - \bar{g}_T(\hat{U}))'\hat{\Omega}(\mu(\theta) - \bar{g}_T(\hat{U})) \tag{A.124}$$

where we recall that

$$\bar{g}_T(\hat{U}) = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^T \hat{u}_t^2 \\ \frac{1}{T}\sum_{t=1}^T \hat{u}_t^4 \\ \frac{1}{T}\sum_{t=1}^T \hat{u}_t^2\hat{u}_{t-1}^2 \end{pmatrix}, \tag{A.125}$$

or else $\bar{g}_T(\hat{U}) = \frac{1}{T}\sum_{t=1}^T g_t(\hat{U})$ and $\hat{u}_t = u_t(\hat{c}_T)$, and $\mu(\theta)' = (\mu_2(\theta), \mu_4(\theta), \mu_{2,2}(1|\theta))$. The first order conditions (F.O.C) associated with this problem are:

$$\frac{\partial \mu'}{\partial \theta}(\hat{\theta}_T)\hat{\Omega}(\mu(\hat{\theta}_T) - \bar{g}_T(\hat{U})) = 0 \,.$$

An expansion of the F.O.C above around the true value θ yields

$$\frac{\partial \mu'}{\partial \theta}(\hat{\theta}_T)\hat{\Omega}\left(\mu(\theta) + \frac{\partial \mu}{\partial \theta'}(\theta)(\hat{\theta}_T - \theta) - \bar{g}_T(\hat{U})\right) \simeq 0$$

after rearranging the equation we have

$$\sqrt{T}(\hat{\theta}_T(\Omega) - \theta) \simeq \left(\frac{\partial \mu'}{\partial \theta}(\theta)\Omega\frac{\partial \mu}{\partial \theta'}(\theta)\right)^{-1}\frac{\partial \mu'}{\partial \theta}(\theta)\Omega\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta)) \,.$$

Using then, propositions **3.5** and **3.7** we get the asymptotic normality of $\hat{\theta}_T(\Omega)$ with asymptotic covariance matrix $W(\Omega)$ as specified in proposition **3.9**.

□

Proof of Proposition 4.1

The proofs derived here follow the lines of Gourieroux,Monfort,Renault (1993). The parameter θ such that $\theta' = (a, r_y, r_w)$ is partitioned into two subvectors

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$

with $\theta_1 = a$ and $\theta_2' = (r_y, r_w)$. The null hypothesis is defined by $H_0 : \{\theta_1 = 0\}$ which corresponds to test the absence of long memory in the model, i.e. $a = 0$.

The expansion given earlier can be rewritten under the null hypothesis with the optimal metric Ω^{*-1}:

$$\sqrt{T}\begin{bmatrix} \hat{\theta}_{1T} \\ \hat{\theta}_{2T} - \theta_2 \end{bmatrix} \sim \left[\begin{pmatrix} \frac{\partial \mu'}{\partial \theta_1} \\ \frac{\partial \mu'}{\partial \theta_2}, \end{pmatrix} \Omega^{*-1}\left(\frac{\partial \mu}{\partial \theta_1'}, \frac{\partial \mu}{\partial \theta_2'}\right)\right]^{-1} \begin{pmatrix} \frac{\partial \mu'}{\partial \theta_1} \\ \frac{\partial \mu'}{\partial \theta_2}, \end{pmatrix} \Omega^{*-1}\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta))$$

$$\sqrt{T}\hat{\theta}_{1T} \simeq (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}(\frac{\partial\mu'}{\partial\theta_1} - A_{12}A_{22}^{-1}\frac{\partial\mu'}{\partial\theta_2})\Omega^{*-1}\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta)) .$$

We note that:

$$\frac{\partial\mu'}{\partial\theta_1} - A_{12}A_{22}^{-1}\frac{\partial\mu'}{\partial\theta_2} = \frac{\partial\mu'}{\partial\theta_1}[Id - M_2]'$$

where

$$M_2 = \frac{\partial\mu}{\partial\theta_2'}\left[\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\frac{\partial\mu}{\partial\theta_2'}\right]^{-1}\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1} \tag{A.126}$$

and $A_{ij} = \frac{\partial\mu'}{\partial\theta_i}\Omega^{*-1}\frac{\partial\mu}{\partial\theta_j}$ which yields

$$\sqrt{T}\hat{\theta}_{1T} \simeq (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}\frac{\partial\mu'}{\partial\theta_1}[Id - M_2]'\Omega^{*-1}\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta_0)) . \tag{A.127}$$

Thus,

$$Var_{as}(\sqrt{T}\hat{\theta}_{1T}) = (A_{11}-A_{12}A_{22}^{-1}A_{21})^{-1}\frac{\partial\mu'}{\partial\theta_1}[Id-M_2]'\Omega^{*-1}[Id-M_2]\frac{\partial\mu}{\partial\theta_1'}(A_{11}-A_{12}A_{22}^{-1}A_{21})^{-1}$$

and we show that $\frac{\partial\mu'}{\partial\theta_1}[Id - M_2]'\Omega^{*-1}[Id - M_2]\frac{\partial\mu}{\partial\theta_1'} = (A_{11} - A_{12}A_{22}^{-1}A_{21})$ yielding

$$W_1 \equiv Var_{as}(\sqrt{T}\hat{\theta}_{1T}) = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} . \tag{A.128}$$

Indeed,

$$\begin{aligned}
\frac{\partial\mu'}{\partial\theta_1}[Id - M_2]'\Omega^{*-1}[Id - M_2]\frac{\partial\mu}{\partial\theta_1'} &= \frac{\partial\mu'}{\partial\theta_1}\left(Id - \frac{\partial\mu}{\partial\theta_2'}\left[\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\frac{\partial\mu}{\partial\theta_2'}\right]^{-1}\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\right)'\Omega^{*-1} \\
&\quad \left(Id - \frac{\partial\mu}{\partial\theta_2'}\left[\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\frac{\partial\mu}{\partial\theta_2'}\right]^{-1}\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\right)\frac{\partial\mu}{\partial\theta_1'} \\
&= \frac{\partial\mu'}{\partial\theta_1}\left(Id - \Omega^{*-1}\frac{\partial\mu}{\partial\theta_2'}\left[\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\frac{\partial\mu}{\partial\theta_2'}\right]^{-1}\frac{\partial\mu'}{\partial\theta_2}\right)\Omega^{*-1} \\
&\quad \left(Id - \frac{\partial\mu}{\partial\theta_2'}\left[\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\frac{\partial\mu}{\partial\theta_2'}\right]^{-1}\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\right)\frac{\partial\mu}{\partial\theta_1'} \\
&= \left(\frac{\partial\mu'}{\partial\theta_1}\Omega^{*-1} - \frac{\partial\mu'}{\partial\theta_1}\Omega^{*-1}\frac{\partial\mu}{\partial\theta_2'}A_{22}^{-1}\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\right) \\
&\quad \left(\frac{\partial\mu}{\partial\theta_1'} - \frac{\partial\mu}{\partial\theta_2'}A_{22}^{-1}\frac{\partial\mu'}{\partial\theta_2}\Omega^{*-1}\frac{\partial\mu}{\partial\theta_1'}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\partial \mu'}{\partial \theta_1}\Omega^{*-1}\frac{\partial \mu}{\partial \theta_1'} - \frac{\partial \mu'}{\partial \theta_1}\Omega^{*-1}\frac{\partial \mu}{\partial \theta_2'}A_{22}^{-1}\frac{\partial \mu'}{\partial \theta_2}\Omega^{*-1}\frac{\partial \mu}{\partial \theta_1'} \\
&= A_{11} - A_{12}A_{22}^{-1}A_{21} \ .
\end{aligned}
$$

Thus, the Wald statistic

$$\xi_T^W = T\hat{\theta}_{1T}'\hat{W}_1^{-1}\hat{\theta}_{1T}$$

is asymptotically equivalent to:

$$
\begin{aligned}
\xi_T^W &= T(\bar{g}_T(\hat{U}) - \mu(\theta))'\Omega^{*-1}[Id - M_2]\frac{\partial \mu}{\partial \theta_1'}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} \\
&\quad (A_{11} - A_{12}A_{22}^{-1}A_{21})\left\{\frac{\partial \mu'}{\partial \theta_1}[Id - M_2]'\Omega^{*-1}[Id - M_2]\frac{\partial \mu}{\partial \theta_1'}\right\}^{-1} \\
&\quad (A_{11} - A_{12}A_{22}^{-1}A_{21})(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}\frac{\partial \mu'}{\partial \theta_1}[Id - M_2]'\Omega^{*-1}(\bar{g}_T(\hat{U}) - \mu(\theta)) \ ,
\end{aligned}
$$

that is

$$
\begin{aligned}
\xi_T^W &= T(\bar{g}_T(\hat{U}) - \mu(\theta))'\Omega^{*-1}[Id - M_2]\frac{\partial \mu}{\partial \theta_1'} \\
&\quad \left\{\frac{\partial \mu'}{\partial \theta_1}[Id - M_2]'\Omega^{*-1}[Id - M_2]\frac{\partial \mu}{\partial \theta_1'}\right\}^{-1}\frac{\partial \mu'}{\partial \theta_1} \\
&\quad [Id - M_2]'\Omega^{*-1}(\bar{g}_T(\hat{U}) - \mu(\theta)) \ .
\end{aligned}
$$

The score statistic is based on the gradient of the objective function with respect to θ_1 evaluated at the constrained estimator $\hat{\theta}_T'^c = (0, \hat{r}_y^c, \hat{r}_w^c)$ i.e.

$$
\begin{aligned}
\mathcal{D}_T &= \frac{\partial \mu'}{\partial \theta_1}(\hat{\theta}_T^c)\Omega^{*-1}(\mu(\hat{\theta}_T^c) - \bar{g}_T(\hat{U})) \\
&\simeq \frac{\partial \mu'}{\partial \theta_1}(\theta)\Omega^{*-1}\left(\mu(\theta) + \frac{\partial \mu}{\partial \theta_2'}(\theta)(\hat{\theta}_{2T}^c - \theta_2) - \bar{g}_T(\hat{U})\right) \\
&\simeq -\frac{\partial \mu'}{\partial \theta_1}(\theta)\Omega^{*-1}\left(\bar{g}_T(\hat{U}) - \mu(\theta_0) - \frac{\partial \mu}{\partial \theta_2'}(\theta)(\hat{\theta}_{2T}^c - \theta_2)\right) \\
&\simeq -\frac{1}{\sqrt{T}}\frac{\partial \mu'}{\partial \theta_1}(\theta)\Omega^{*-1}\left(\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta)) - \frac{\partial \mu}{\partial \theta_2'}(\theta)\sqrt{T}(\hat{\theta}_{2T}^c - \theta_2)\right)
\end{aligned}
$$

Given that

$$\sqrt{T}(\hat{\theta}_{2T}^c - \theta_2) \simeq \left(\frac{\partial \mu'}{\partial \theta_2}(\theta)\Omega^{*-1}\frac{\partial \mu}{\partial \theta_2'}(\theta)\right)^{-1}\frac{\partial \mu'}{\partial \theta_2}(\theta)\Omega^{*-1}\sqrt{T}(\bar{g}_T(\hat{U}) - \mu(\theta))$$

we have

$$
\begin{aligned}
\mathcal{D}_T &\simeq -\frac{1}{\sqrt{T}}\frac{\partial \mu'}{\partial \theta_1}(\theta)\Omega^{*-1}\left(\sqrt{T}(\bar{g}_T(\hat{U})-\mu(\theta)) - \frac{\partial \mu}{\partial \theta_2'}(\theta)A_{22}^{-1}\frac{\partial \mu'}{\partial \theta_2}(\theta)\Omega^{*-1}\sqrt{T}(\bar{g}_T(\hat{U})-\mu(\theta))\right) \\
&\simeq -\frac{1}{\sqrt{T}}\frac{\partial \mu'}{\partial \theta_1}(\theta)(Id-M_2)'\Omega^{*-1}\sqrt{T}(\bar{g}_T(\hat{U})-\mu(\theta))
\end{aligned}
$$

where M_2 has been defined at equation (A.126). Finally, from equation (A.127) we have

$$
\mathcal{D}_T \simeq -(A_{11} - A_{12}A_{22}^{-1}A_{21})\hat{\theta}_{1T} \ . \tag{A.129}
$$

There is asymptotically a one-to-one linear relationship between $\mathcal{D}_T$ and the unrestricted estimator $\hat{\theta}_{1T}$ and this shows that the score test is asymptotically equivalent to the Wald test and

$$
Var_{as}(\mathcal{D}_T) = A_{11} - A_{12}A_{22}^{-1}A_{21} \ . \tag{A.130}
$$

On the other hand, the difference between the two optimal values of the objective function (the constrained minus the unconstrained one) is such that

$$
\begin{aligned}
\xi_T^C \simeq\ & T(\bar{g}_T(\hat{U})-\mu(\theta))'[Id-M_2]'\Omega^{*-1}[Id-M_2](\bar{g}_T(\hat{U})-\mu(\theta)) \\
& -(\bar{g}_T(\hat{U})-\mu(\theta))'[Id-M]'\Omega^{*-1}[Id-M](\bar{g}_T(\hat{U})-\mu(\theta))
\end{aligned}
$$

where $M_2 = \frac{\partial \mu}{\partial \theta_2'}A_{22}^{-1}\frac{\partial \mu'}{\partial \theta_2}\Omega^{*-1}$ and $M = \frac{\partial \mu}{\partial \theta_2'}\left(\frac{\partial \mu'}{\partial \theta}\Omega^{*-1}\frac{\partial \mu}{\partial \theta'}\right)^{-1}\frac{\partial \mu'}{\partial \theta}\Omega^{*-1}$. Thus,

$$
\begin{aligned}
\xi_T^C &\simeq T(\bar{g}_T(\hat{U})-\mu(\theta))'\Omega^{*-1}[Id-M_2](\bar{g}_T(\hat{U})-\mu(\theta)) \\
&\quad -(\bar{g}_T(\hat{U})-\mu(\theta))'\Omega^{*-1}[Id-M](\bar{g}_T(\hat{U})-\mu(\theta)) \\
&\simeq T(\bar{g}_T(\hat{U})-\mu(\theta))'\Omega^{*-1}[M-M_2](\bar{g}_T(\hat{U})-\mu(\theta)) \ .
\end{aligned}
$$

A classical argument of block inverse gives

$$
\begin{aligned}
\Omega^{*-1}[M-M_2] &= \Omega^{*-1}[Id-M_2]\frac{\partial \mu}{\partial \theta_1'}\left(\frac{\partial \mu'}{\partial \theta_1}[Id-M_2]'\Omega^{*-1}[Id-M_2]\frac{\partial \mu}{\partial \theta_1'}\right)^{-1} \\
&\quad \frac{\partial \mu'}{\partial \theta_1}[Id-M_2]'\Omega^{*-1}
\end{aligned}
$$

and the asymptotic equivalence between ξ_T^C and ξ_T^W follows. □

References

ANDERSEN, T., G. (1994): "Stochastic Autoregressive Volatility: A Framework for Volatility Modelling.," *Mathematical Finance*, 4(2), 75–102.

ANDERSEN, T., AND B. SORENSEN (1996): "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Economics and Business Statistics*, 14(3), 328–352.

ANDERSEN, T. G., H.-J. CHUNG, AND B. E. SØRENSEN (1999): "Efficient Method of Moments Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Econometrics*, 91, 61–87.

ANDREWS, D. (1987): "Asymptotic Results for Generalized Wald Tests," *Econometric Theory*, 3, 348–358.

BAILLIE, R., AND H. CHUNG (2001): "Estimation of GARCH Models from the Autocorrelations of the Squares of a Process," *Journal of Time Series mAnalysis*, 22(6), 631–650.

BANSAL, R., A. R. GALLANT, R. HUSSEY, AND G. E. TAUCHEN (1995): "Nonparametric Estimation of Structural Models for High-Frequency Currency Market Data," *Journal of Econometrics*, 66, 251–287.

BARNARD, G. A. (1963): "Comment on "The Spectral Analysis of Point Processes" by M. S. Bartlett," *Journal of the Royal Statistical Society, Series B*, 25, 294.

BERGER, A., AND S. WALLENSTEIN (1989): "On the Theory of $C(\alpha)$ tests," *Statistics and Probability Lettres*, 7, 419–424.

BIRNBAUM, Z. W. (1974): "Computers and Unconventional Test-Statistics," in *Reliability and Biometry*, ed. by F. Proschan, and R. J. Serfling, pp. 441–458. SIAM, Philadelphia, PA.

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics*, 51, 307–327.

CARRASCO, M., AND X. CHEN (1999): "β-mixing and Moment Properties of Various GARCH, Stochastic Volatility and ACD Models," Discussion paper, London School of Economics.

CHEN, M., AND G. CHEN (2000): "Geometric ergodicity of nonlinear autoregressive models with changing conditional variances," *The Canadian Journal of Statistics*, 28(3), 605–613.

Danielsson, J. (1994): "Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood," *Journal of Econometrics*, 61, 375–400.

Danielsson, J., and J.-F. Richard (1993): "Accelerated Gaussian Importance Sampler with Application to Dynamic Latent Variable Models," *Journal of Applied Econometrics*, 8, S153–S173.

Davidson, J. (1994): *Stochastic Limit Theory*, Advanced Texts in Econometrics. Oxford University Press Inc., New York.

Duffie, D., and K. J. Singleton (1993): "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–952.

Dufour, J.-M. (1995): "Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics in Econometrics," Discussion paper, C.R.D.E., Université de Montréal.

Dufour, J.-M., L. Khalaf, J.-T. Bernard, and I. Genest (2001): "Simulation-Based Finite-Sample Tests for Heteroscedasticity and ARCH Effects," Discussion Paper 08-2001, Universitè de Montrèal, C.R.D.E.

Dufour, J.-M., and A. Trognon (2001): "Invariant tests based on M-estimators, estimating functions, and the generalized method of moments," Discussion paper, CIREQ, University of Montreal.

Dwass, M. (1957): "Modified Randomization Tests for Nonparametric Hypotheses," *Annals of Mathematical Statistics*, 28, 181–187.

Engle, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

Foutz, R. V. (1980): "A Method for Constructing Exact Tests from Test Statistics that have Unknown Null Distributions," *Journal of Statistical Computation and Simulation*, 10, 187–193.

Fuller, W. A. (1976): *Introduction to statistical Time Series*. John Wiley and Sons, New York.

Gallant, A., D. Hsieh, and G. Tauchen (1995): "Estimation of Stochastic Volatility Models with Diagnostics," Discussion paper, Duke University.

Gallant, A. R., and G. Tauchen (1996): "Which Moments to Match?," *Econometric Theory*, 12, 657–681.

Ghysels, E., A. Harvey, and E. Renault (1996): "Stochastic Volatility," in *Statistical Methods in Finance*, ed. by C. Rao, and G. Maddala, Holland. North-Holland.

Goffe, W. L., G. D. Ferrier, and J. Rogers (1994): "Global Optimization of Statistical Functions with Simulated Annealing," *Journal of Econometrics*, 60, 65–99.

Gouriéroux, C. (1997): *ARCH Models and Financial Applications*, Springer Series in Statistics. Springer-Verlag, New York.

Gouriéroux, C., and A. Monfort (1995): *Simulation Based Econometric Methods*, CORE Lecture Series. CORE Foundation, Louvain-la-Neuve.

Gouriéroux, C., A. Monfort, and E. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8S, 85–118.

Gradshteyn, I., and I. Ryzhik (1980): *Table of integrals, series, and products. Corrected and enlarged edition.* Academic Press inc.

Hamilton, J. (1994): *Time Series Analysis.* Princeton University Press Inc., New Jersey.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Harvey, A., E. Ruiz, and N. Shephard (1994): "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247–264.

Harvey, A., and N. Shephard (1996): "Estimation of an asymmetric stochastic volatility model for asset returns," *Journal of business and Economic statistics*, 14, 429–434.

Jacquier, E., N. Polson, and P. Rossi (1994): "Bayesian Analysis of Stochastic Volatility Models (with discussion)," *Journal of Economics and Business Statistics*, 12, 371–417.

Kim, S., and N. Shephard (1994): "Stochastic Volatility: Optimal Likelihood Inference and Comparison with ARCH Models," Discussion paper, Nuffield College, Oxford, unpublished paper.

Kim, S., N. Shephard, and S. Chib (1996): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," Discussion paper, Nuffield College, Oxford, unpublished paper.

——— (1998): "Stochastic volatility : Likelihood inference and comparison with ARCH models," *Review of economic studies*, 65, 361–393.

Kocherlakota, S., and K. Kocherlakota (1991): "Neyman's $C(\alpha)$ test and Rao's efficient score test for composite hypotheses," *Statistics and Probability Lettres*, 11, 491–493.

Loève, M. (1963): *Probability Theory*. Van Nostrand Reinhold, LTD., third edn.

Mahieu, R., and P. Schotman (1998): "An Empirical application of stochastic volatility models," *Journal of Applied Econometrics*, 13, 333–360.

Marriott, F. H. C. (1979): "Barnard's Monte Carlo Tests: How Many Simulations?," *Applied Statistics*, 28, 75–77.

Melino, A., and S. Turnbull (1990): "Pricing Foreign Currency Options with Stochastic Volatility," *Journal of Econometrics*, 45, 239–265.

Monfardini, C. (1997): "Estimating Stochastic volatility Models through Indirect Inference," Discussion paper, European University Institute.

Nelson, D. (1988): "Times series behavior of stock market volatility and returns," unpublished Ph.P dissertation, Massachusetts Institute of technology, Economics Dept.

Newey, W. K., and K. D. West (1987a): "Hypothesis Testing with Efficient Method of Moments Estimators," *International Economic Review*, 28, 777–787.

——— (1987b): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Ronchetti, E. (1987): "Robust $C(\alpha)$-type tests for linear models," *Sankhya: The Indian journal of Statistics*, 49(Series A), 1–16.

Rudin, W. (1976): *Principles of Mathematical Analysis*. McGraw-Hill, Inc.

Ruiz, E. (1994): "Quasi-maximum likelihood estimation of stochastic variance models," *Journal of Econometrics*, 63, 284–306.

Shephard, N. (1996): "Statistical Aspects of ARCH and Stochastic Volatility," in *Time Series Models in Econometrics, Finance and Other Fields*, ed. by D. Cox, O. Barndorff-Nielsen, and D. Hinkley. Chapman & Hall, London.

Tauchen, G. (1996): "New Minimum Chi-Square Methods in Empirical Finance," Discussion paper, Department of Economics, Duke University.

Taylor, S. (1986): *Modelling Financial Time Series*. John Wiley, Chichester.

——— (1994): "Modelling Stochastic Volatility," *Mathematical Finance*, 4, 183–204.

Wong, C. (2002a): "Estimating stochastic volatility:new method and comparison," Discussion paper, University of Oxford, Ph.d. Thesis.

——— (2002b): "The MCMC method for an extended stochastic volatility model," Discussion paper, University of Oxford, Ph.d. Thesis.